DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS

Mohamed Abdelnasser Mohamed Mahmoud

Development of advanced computer methods for breast cancer image interpretation through texture and temporal evolution analysis

DOCTORAL THESIS

Author:
Mohamed Abdel-Nasser Mohamed Mahmoud

Advisors:
Dr. Domènec Savi Puig Valls
Dr. Antonio Moreno Ribas

Departament d'Enginyeria Informàtica i Matemàtiques

UNIVERSITAT ROVIRA I VIRGILI
Tarragona
2016

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

UNIVERSITAT ROVIRA I VIRGILI
Departament d'Enginyeria Informàtica i Matemàtiques

Av. Paisos Catalans, 27
43007 Tarragona
Tel. +34 977 55 95 95
Fax. +34 977 55 95 97

We STATE that the present study, entitled "Development of advanced computer methods for breast cancer image interpretation through texture and temporal evolution analysis", presented by Mohamed Abdel-Nasser Mohamed Mahmoud, for the award of the degree of Doctor, has been carried out under our supervision at the Departament d'Enginyeria Informàtica i Matemàtiques.

Tarragona, May 2016.

Doctoral Thesis Supervisors,

Dr. Domènec Savi Puig Valls

Dr. Antonio Moreno Ribas

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

To my daughter Rafef, my wife Nada, my sister and my mother

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

Abstract

Breast cancer is one of the most dangerous diseases that attack women. Computer-aided diagnosis systems may help to detect breast cancer early and reduce mortality. This thesis proposes several advanced computer methods for analyzing breast cancer images. We analyze breast cancer in three imaging modalities: mammography, ultrasonography and thermography. Our analysis includes mass/normal breast tissue classification, benign/malignant tumor classification in mammograms and ultrasound images, nipple detection in thermograms, mammogram image registration and analysis of the evolution of breast tumors.

We studied the performance of various texture analysis methods so that the number of false positives in breast cancer detection could be reduced. We considered such well-known texture analysis methods as local binary patterns, histogram of oriented gradients, co-occurrence matrix features and Gabor filters, and proposed two texture descriptors: uniform local directional pattern and fuzzy local directional pattern. We also studied the effect of factors such as pixel resolution, integration scale, preprocessing and feature normalization on the performance of these texture methods for tumor classification. Finally, we used super-resolution approaches to improve the performance of texture analysis methods when classifying breast tumors in ultrasound images. The methods proposed discriminated between different tissues, and significantly improve the analysis of breast cancer.

For the analysis of breast cancer in thermograms, we propose an unsupervised, automatic method for detecting nipples that is accurate, simple and fast. To analyze the evolution of breast cancer, we propose a temporal mammogram registration method based on the curvilinear coordinates. We also propose a method for quantifying and visualizing the evolution of breast tumors in patients undergoing medical treatment that uses flow fields, ordered weighted averaging aggregation operators and strain tensors. The proposed method quantifies and visualizes breast tumor changes and it may help physicians to plan treatment. Overall, the methods proposed in this thesis improve the performance of the state-of-the-art approaches

and may help to improve the diagnosis of breast cancer.

Keywords: Breast Cancer, Benign, Malignant, Mammogram, Ultrasound, Thermography, Computer-aided Diagnosis System, Feature Extraction, Texture Methods, Classification, Super-Resolution, Registration, Optical Flow.

Acknowledgements

The author was supported by a PhD grant from URV in 2013 (Martí Franquès program).

This work was partly supported by the Spanish Government through project TIN2012-37171-C02-02.

I would like to express my gratitude to my supervisors Dr. Antonio Moreno and Dr. Domenec Puig for their useful guidance, insightful comments, and considerable encouragement to complete this thesis. I also thank Jaime Melendez and Hatem Rashwan for their cooperation and support. Special thanks to Julián Cristiano for giving me the right to use the Latex and cover templates of his thesis, besides several useful comments in the last three years. I also want to thank Adel Saleh, Kerlos Atia, Mohamed Jebreel and Farhan Akram for their support and encouragement.

Contents

Contents xi

List of Figures

List of Tables

Part I

Introduction

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

CHAPTER 1

Introduction

1.1 Motivation

Breast cancer is one of the most dangerous diseases because it attacks women in their 40s worldwide. According to European Union statistics, breast cancer is the leading cause of cancer death (Malvezzi et al., 2015), and 90800 breast cancer deaths are predicted in 2015 (14.22 mortality rate for 2015 vs. 15.85 for 2009 per 100,000 population). Death rates from breast cancer have decreased since 1989, with a big reduction in women younger than 50. This reduction is a result of earlier detection through screening with computer-aided diagnosis (CAD) systems. Many countries (e.g. Spain and the UK) have established regular screening programs for their women in an attempt to detect the early signs of breast cancer, in which the screening is usually performed every two years.

Physicians usually order mammograms (X-ray images of the breast) to help in their

diagnosis of breast cancer, especially to detect tumours that cannot be easily felt. The common risk factors of breast cancer are *age*, *family profile*, *genetics* and *breast density*. Breast density, which represents the amount of dense tissue in the breast, is regarded as one of the strongest risk factors of breast cancer (Lokate et al., 2010). The higher the breast density is, the higher the probability of developing breast cancer.

Mammograms are considered to be the best method for early detection of breast cancer. However, even though a breast mammogram may show suspicious regions, it cannot in itself prove that an abnormal area is cancerous. If a mammogram presents a suspicion of cancer, a biopsy should be performed. This involves extracting a sample from a suspicious breast tissue which is then analyzed under a microscope. Biopsies are expensive (1,000 to 5,000 dollars[1]), painful and a psychological burden for the patient. Thus, a reduction in the number of false positives (mammograms interpreted as abnormal by a CAD system when they are actually normal) would yield a huge reduction in the cost of diagnosing breast cancer. In practice, several methods have been used as an adjunct to mammography, such as ultrasound, magnetic resonance and infrared images.

In the last three decades, various CAD systems have been proposed for automatic breast mass detection using computer vision and machine learning techniques and the advances in breast cancer diagnosis and CAD systems have led to a $30-50\%$ fall in mortality in several countries. These methods can be classified as *supervised* or *unsupervised*. In the supervised methods, which are more accurate, feature vectors are extracted from the regions of interest (ROIs) . The extracted features can be related to *texture*, *statistic*, *position* and *geometry*. Fig. 1.1 presents the main steps in the breast cancer CAD system: *ROIs extraction*, *feature extraction* and *classification*. The normal and mass regions represent the most interesting ROIs in breast images. The extracted features are used to train a model, which is used to predict the class of unknown instances. Unfortunately, supervised methods usually present a high number of false positive detections. In addition, physicians have few

[1] http://health.costhelper.com/biopsy.html

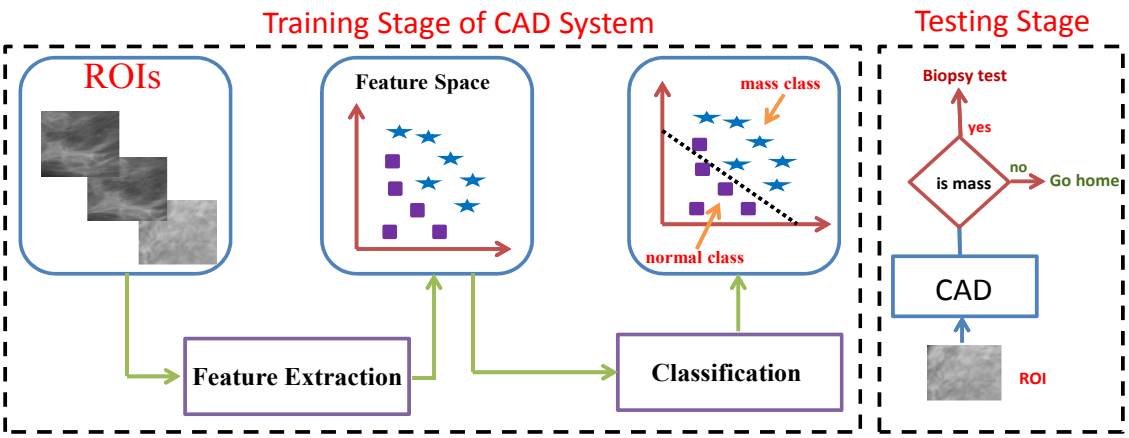


Figure 1.1: The main steps in a breast cancer CAD system

computer programs available to help them predict the pathological response and adjust the medical treatment so that it has the intended effects. For these reasons, this PhD thesis focuses on making a useful contribution to improving the performance of breast cancer CAD systems. Our research focuses on analyzing breast cancer in mammograms, ultrasound and infrared images, and we study the evolution of breast tumors using image registration and optical flow methods.

1.2 Thesis objectives

The main objectives of this thesis are:

1. To analyze breast cancer in mammograms. We classify breast tissue into normal or abnormal, and breast tumors into benign or malignant, and estimate breast density. To carry out these studies, we analyze the performance of several CAD systems, feature extraction and classification methods.
2. To evaluate and improve the performance of texture analysis methods in classifying breast tissues in ultrasound images (normal/abnormal and benign/malignant).
3. To propose CAD systems to analyze breast tissue in infrared images (thermograms).
4. To analyze the evolution of breast tumors in temporal mammograms using motion analysis methods.

1.3 Contributions

The main contributions of this thesis are the following:

1. We analyze the performance of several texture analysis method in classifying breast tissues in mammograms into normal or abnormal. We propose several combinations of texture analysis methods and classifiers to improve the classification results. We also propose a new texture analysis method, called uniform local directional pattern (ULDP) to analyze the abnormality of breast tissues and estimate breast tissue density. Moreover, we analyze the impact of pixel resolution, patch size, preprocessing and feature normalization on texture analysis when classifying tumors into benign or malignant.

   The results of the previous studies have been published in the following journals:

   - Mohamed Abdel-Nasser, Hatem A. Rashwan, Domenec Puig, Antonio Moreno, "Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern," *Experts Systems with Applications*, 42(24): 9499-9511 (2015). Impact Factor: 2.24 (Q1).
   - Mohamed Abdel-Nasser, Antonio Moreno, Domenec Puig, "Towards Cost Reduction of Breast Cancer Diagnosis using Mammography Texture Analysis," *Journal of Experimental & Theoretical Artificial Intelligence*, 28(1-2): 385-402 (2016). Impact Factor: 1.00 (Q3).
   - Mohamed Abdel-Nasser, Jaime Melendez, Antonio Moreno, Domenec Puig, "The impact of pixel resolution, integration scale, preprocessing and feature normalization on texture analysis for mass classification in mammograms," *International Journal of Optics* (in press). SCImago Journal Rank: 0.3 (Q3).

   Preliminary work on improving the performance of texture method for mass detection in mammograms has been presented in the following book chapter:

   - Mohamed Abdel-Nasser, Domenec Puig, Antonio Moreno, "Improvement

of Mass Detection In Breast X-Ray Images Using Texture Analysis Methods," *Artificial Intelligence Research and Development: Recent Advances and Applications*, vol. 269, pp. 159-168, IOS press, 2014.

2. We propose a fuzzy version of ULDP, called fuzzy local directional pattern (FLDP) to cope with the noise and fuzzy regions that may appear in ultrasound images and mammograms. We also propose the use of image super-resolution to improve the performance of texture analysis methods when classifying breast tumors in ultrasound images. We show that our super-resolution-based approach improves the performance of texture methods and thus outperforms the state of the art in benign/malignant tumor classification.

   The results of the previous study are under review in the following journal:

   - Mohamed Abdel-Nasser, Jaime Melendez, Antonio Moreno, Osama A. Omer, Domenec Puig, "Breast tumor classification in ultrasound images using texture analysis and super-resolution methods," *Engineering Applications of Artificial Intelligence*. Impact factor: 2.21 (Q1).

   Preliminary work on improving the performance of the texture method with ultrasound images has been presented at the following international conference:

   - Mohamed Abdel-Nasser, Domenec Puig, Antonio Moreno, Adel Saleh, Joan Marti, Luis Martin, Anna Magarolas, "Breast Tissue Characterization in X-Ray and Ultrasound Images using Fuzzy Local Directional Patterns and Support Vector Machines," *Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISAPP2015)*, Vol. 1, pp. 387-394, 2015. Core C.

3. We propose an automatic, accurate and fast method for detecting nipples in thermograms. The main stages in the method proposed are: human body segmentation, determination of nipple candidates using adaptive thresholding and detecting the nipples using a proposed nipple selection algorithm.

   The results of the previous study are under review in the following journal:

   - Mohamed Abdel-Nasser, Adel Saleh, Antonio Moreno, Domenec Puig, "Automatic nipple detection in thermograms using image processing and

anatomical information," *Experts Systems with Applications*. Impact Factor: 2.24 (Q1).

4. A temporal mammogram registration method is proposed in this thesis. It is based on the curvilinear coordinates, which are used to cope with both global and local deformations in the breast area.

   The results of the previous study have been published in the following journal:

   - Mohamed Abdel-Nasser, Antonio Moreno, Domenec Puig, "Temporal mammogram image registration using optimized curvilinear coordinates," *Computer Methods and Programs in Biomedicine*, 127, pages 1-14 (2016). Impact factor: 1.9 (Q1).

5. We propose a method for quantifying and visualizing the evolution of breast tumors in patients undergoing medical treatments with flow fields and strain tensors. The method determines the displacement fields between each follow-up mammogram and its baseline. The resulting displacement fields are then used to calculate the strain tensors.

   The results of the previous study have been published in the following journal:

   - Mohamed Abdel-Nasser, Hatem A. Rashwan, Antonio Moreno, Luis Martin, Meritxell Arenas, Anna Magarolas, Lorena Diez-Presa, Joan Marti, Domenec Puig, "Breast Cancer Development Analysis In Follow-Up Digital Mammograms Through Anatomical-Based Variational Optical Flow: Preliminary Study," *International Journal of Computer Assisted Radiology and Surgery*, Vol. 10 (Suppl 1): S1-S312. Impact factor: 1.71 (Q2).
   - Mohamed Abdel-Nasser, Antonio Moreno, Hatem A. Raswan, Domenec Puig, "Analyzing the evolution of breast tumors using flow fields and strain tensors," *Pattern Recognition Letters* (under review). Impact factor: 1.55 (Q2).

   Preliminary work on this topic has been presented in the following book chapter:

   - Mohamed Abdel-Nasser, Antonio Moreno, Domenec Puig, "Analysis

of the evolution of breast tumours using strain tensors," *Artificial Intelligence Research and Development: Proceedings of the 18th International Conference of the Catalan Association for Artificial Intelligence*, vol. 277, pp. 237-246, IOS press, 2015.

1.4 Thesis organization

The thesis is divided into the following six parts:

- Part I: Introduction
  - Chapter 1: *Introduction*

    This chapter introduces breast cancer CAD systems. It starts with the motivation behind the thesis and the contributions it makes to improving cancer CAD systems.
  - Chapter 2: *Background*

    In this chapter we describe the background to breast cancer screening methods, the stages in the breast cancer CAD system, CAD evaluation, motion analysis methods and breast cancer databases.
- Part II: Analysis of breast cancer in mammograms
  - Chapter 3: *Mass analysis in mammograms using texture methods*

    In this chapter we compare several texture analysis methods for breast mass detection. To improve mass detection rates, we propose using two combination schemes. Firstly, we concatenate the best texture analysis methods. Secondly, we use the classifier voting technique to combine the predictions given by the best methods.
  - Chapter 4: *Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern*

    This chapter proposes a CAD system to classify breast tissue into normal or mass, and to estimate the breast density. We propose the ULDP descriptor for feature extraction. ULDP encodes a neighborhood in the breast region based on its edge responses, in addition to spatial

information.

  - Chapter 5: *The impact of image resolution, patch size, preprocessing and feature normalization on texture analysis for tumor classification*

    In this chapter we study the impact of factors such as pixel resolution, integration scale, preprocessing and feature normalization on the performance of texture methods when applied to tumor classification.

- Part III: Analysis of breast cancer in ultrasound images
  - Chapter 6: *Breast tissue characterization in ultrasound images using fuzzy local directional patterns*

    In this chapter we propose the fuzzy local directional patterns (FLDP) for breast tissue characterization. FLDP describes each pixel in a given image by its edge responses and makes use of fuzzy membership functions.
  - Chapter 7: *Breast tumor classification in ultrasound images using texture analysis and super-resolution methods*

    To improve the performance of texture methods when applied to tumor classification in ultrasound images, this chapter proposes the use of a super-resolution approach that exploits the complementary information provided by multiple images of the same target. The proposed CAD system consists of four stages: super-resolution computation, extraction of the region of interest (ROI), feature extraction and classification.
- Part IV: Analysis of breast cancer in infrared images
  - Chapter 8: *Automatic nipple detection in thermograms using image processing and anatomical information*

    In this chapter we propose an automatic, accurate and real-time method to detect nipples in infrared images (thermograms). The main stages of the method are: human body segmentation, determination of nipple candidates using adaptive thresholding and finally nipple detection using rules derived from the anatomical structure of the human body that appear in thermograms.
- Part V: Analysis of temporal evolution of breast cancer

  - Chapter 9: *Temporal mammogram image registration using optimized curvilinear coordinates*

    This chapter proposes a registration method for aligning temporal mammograms. It uses a curvilinear coordinate system to align the mammograms. In this way the system uses anatomical-driven coordinates instead of Cartesian coordinates, which ignore the anatomical structure of the breast.
  - Chapter 10: *Analyzing the evolution of breast tumors using flow fields and strain tensors*

    In this chapter we propose a method for quantifying and visualizing the changes in breast tumors in patients undergoing medical treatment through flow fields and strain. It determines the displacement fields between each follow-up mammogram and its baseline. The resulting displacement fields are then used to calculate the strain tensors.

- Part VI: *Conclusion*
  - Chapter 11: Concluding remarks.

    This chapter presents the conclusions of the thesis and some lines of future research.

CHAPTER 2

Background

2.1 Introduction

Several screening methods have been used for the early detection of breast cancer, for example, mammography, ultrasonography and magnetic resonance imaging. Several CAD systems have been proposed for analyzing breast cancer images using computer vision, pattern recognition and machine learning techniques. CAD systems that classify breast tissues into normal/abnormal or benign/malignant usually consist of three main stages: segmentation of the region of interest (ROI), feature extraction and classification. In this chapter, we present the background to breast cancer screening methods, the stages in the breast cancer CAD system, CAD evaluation and motion analysis methods.

2.2 Breast cancer screening

Breast cancer screening is the process of checking women's breasts for cancer before there are signs or symptoms of the disease. A breast mainly consists of lobules, ducts and stroma. Lobules are milk producing glands, ducts are small pipes that carry the milk from the lobules to the nipple, and stroma is a fatty and connective tissue surrounding the ducts and lobules, blood vessels, and lymphatic vessels. Physicians rely on mammograms to diagnose breast cancer, especially to detect tumors that cannot be easily felt. The following subsections present the common breast cancer screening methods.

2.2.1 Mammography

Mammograms are X-ray images of the breast. The common screening mammographic views are cranio-caudal (CC) and medioLateral oblique (MLO) (Elmore et al., 2005). As shown in Fig. 2.1, a CC mammographic view is captured from above of a horizontally compressed breast (the breast is compressed using compression plates). In turn, the MLO is taken from the side and at an angle of a diagonally compressed breast[1].

Tomosynthesis is a 3D mammography in which a machine takes many low-dose X-rays as it moves over the breast. The images taken can be combined into a three-dimensional picture, which may allow doctors to see inside the breast more clearly than with a standard 2D mammogram, and possibly detect more masses.

2.2.2 Breast ultrasonography (BUS)

Ultrasonography is an imaging method that transmits high-frequency sound waves through the breast and converts them into images. There is no radiation involved during the acquisition of the images. Mammograms may fail to detect cancer in young women because their breasts tend to be dense. In these cases, ultrasonography may provide help. However, ultrasound is usually used to

[1] http://en.wikiversity.org

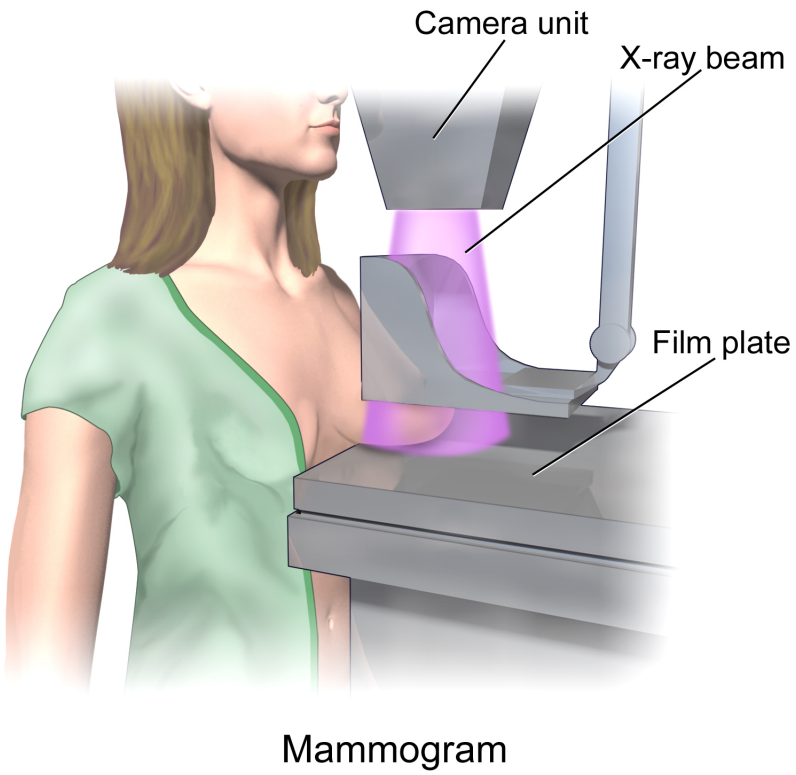


Figure 2.1: Acquisition of a mammogram (Blausen Gallery 2014)

complement mammography. If an abnormality is seen on mammography or felt by physical exam, ultrasound is the best way to find out if the abnormality is solid (such as a benign fibroadenoma or cancer) or fluid-filled (such as a benign cyst). It can not determine whether a solid lump is cancerous or detect calcifications.

The automated breast ultrasound system (ABUS) is a comfortable, non-ionizing alternative to other supplemental screening options for women with dense breast tissue. When ABUS is used in addition to mammography, breast cancer detection can be improved by 55% over mammography.

2.2.3 Thermography

Thermography uses a special camera to measure the temperature of the skin on the surface of the breast. It is a non-invasive method that involves no radiation. Thermography is based on two principles: 1) cancer cells grow and multiply very fast, so blood flow and metabolism are higher in regions that contain a tumor than

in other regions, 2) skin temperature goes up as blood flow and metabolism increase. Indeed, thermography may be used to supplement information from a mammogram and help identify cancers that are close to the skin. In practice, thermography can not find cancers that are deeper in the breast or detect small cancers.

2.2.4 Magnetic resonance imaging (MRI)

MRI is a technology that uses magnets and radio waves to produce detailed cross-sectional images of the inside of the body. MRI does not use X-rays, so it does not involve any exposure to radiation. Breast MRI has a number of different uses for breast cancer, such as screening high-risk women, and gathering more information about an area of suspicion found on a mammogram or BUS images.

2.3 CAD system

A breast cancer CAD system usually consists of three main stages: segmentation of the ROI from the breast image, feature extraction from the ROI, and classification of the ROI.

2.3.1 ROI segmentation

The most important regions of interest (ROIs) in the breast are the ones that contain suspicious tissues. Several methods have been proposed to extract the ROIs from breast images. The simplest method is to extract the ROIs manually. However, there are also many automatic methods. A number of works have used the region growing, thresholding, edge detection segmentation methods to extract the ROI from the breast (Singh and Al-Mansoori, 2000). For instance, Hong and Sohn (2010) proposed an approach to automatically segment the ROIs from mammograms by analyzing both the topological and the geometrical structure of the image based on an isocontour map that effectively provides image features. The authors assumed that the ROIs are salient regions, since they stand out against the surrounding background. Fig. 2.2 shows an example of the ground truth (GT) of a mass region

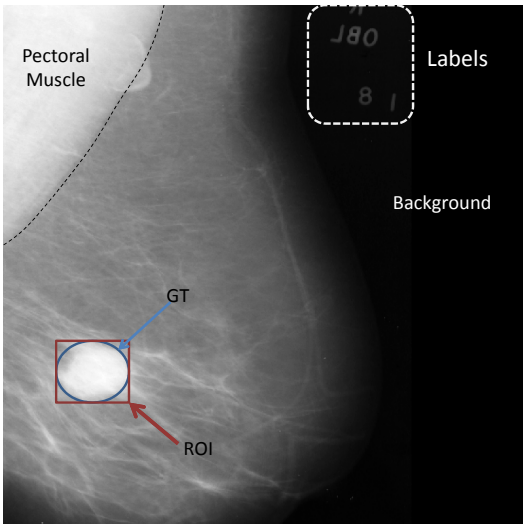


Figure 2.2: ROI extraction from a mammogram

(blue circle), and the ROI (red square surrounding the circle). In turn, the ROIs of the normal tissues can be randomly selected from the normal regions. In the case of BUS images, Cai et al. (2015) manually extracted the ROIs while Gómez et al. (2012) used a semi-automatic ROI extraction method called marker-controlled watershed transformation (Gómez et al., 2010).

2.3.2 Feature extraction

Features related to *texture*, *statistic*, *position* and *geometry* have been used in breast cancer CAD systems. In this thesis, we have considered texture analysis methods which have usually been applied in the field, such as local binary pattern (LBP), histogram of oriented gradients (HOG), grey level co-occurrence matrix (GLCM) features and Gabor filters.

2.3.2.1 Local binary pattern (LBP)

LBP is a grey scale invariant texture feature that is regarded as a good method for texture image analysis in many computer vision areas (Ojala et al., 2002). The original LBP operator labels the pixels of an image by comparing the 3×3 window surrounding each pixel with the value of the central pixel. Pixels in this window with a value greater than the central pixel are labeled 1 and the rest as 0; thus, each

pixel is represented by 8 bits (see Fig. 2.3). The size of the window may vary on different applications (e.g. 3×3, 5×5 or 7×7).

Uniform LBP is an extension of the original LBP in which only patterns that contain at most two transitions from 0 to 1 (or viceversa) are considered. For example *00111100* is a uniform LBP whereas *00101010* is a non-uniform LBP. In uniform LBP mapping there is a separate output label for each uniform pattern and all the non-uniform patterns are assigned to a single label (the uniform mapping produces 59 output labels for neighbourhoods of 8 points).

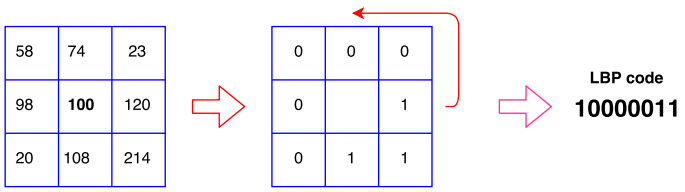


Figure 2.3: Calculation of LBP for a given pixel in a small neighborhood

2.3.2.2 Histogram of oriented gradients (HOG)

HOG has been considered as a robust feature extraction method because it produces distinctive features in the case of illumination change and cluttered background (Dalal and Triggs, 2005). In the HOG method, the occurrences of edge orientations in a local image window are counted. The image is divided into blocks (small groups of cells) and then a weighted histogram is computed for each of them (see Fig. 2.4). The frequencies in the histograms are normalized in the interval [0,1] to compensate changes in illumination. The combination of the histograms of all those blocks represents the HOG descriptors.

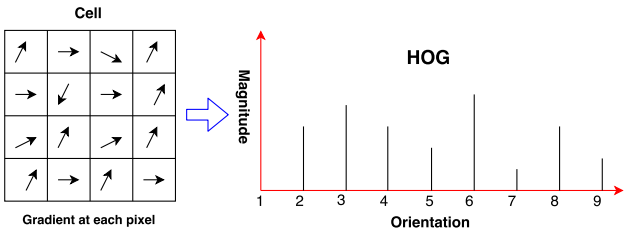


Figure 2.4: Calculation of HOG descriptor

2.3.2.3 Gray level co-occurrence matrix (GLCM)

In the GLCM the distribution of co-occurring grey level values in a given direction and at a given distance is computed (Haralick et al., 1973). In other words, the GLCM computes the joint frequencies $p(i, j)$ of pairwise combinations of gray levels i and j separated by distance d along direction θ (see Fig. 2.5). If N_g is the number of distinct gray levels in the quantized image, then the size of the GLCM is N_g^2. The GLCM features and their mathematical expressions are listed in Table 2.1 (Haralick et al., 1973; Soh and Tsatsoulis, 1999; Clausi, 2002). In addition, we present some terms used to compute the GLCM features in Table 2.2 (Haralick et al., 1973; Soh and Tsatsoulis, 1999).

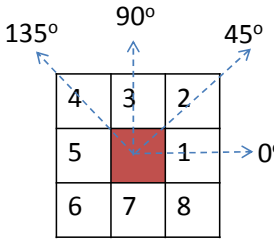


Figure 2.5: Calculation of GLCM along different directions

2.3.2.4 Gabor filters

A two dimensional Gabor filter $g(x, y)$ can be viewed as a sinusoid with a particular frequency and orientation, modulated by a Gaussian envelope

$$g(x,y) = exp^{\frac{-1}{2}(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2})} exp^{-j2\pi(u_0x+v_0y)} \tag{2.1}$$

where (u_0, v_0) is the centre of a sinusoidal function and σ_x, σ_y are the standard deviations along two orthogonal directions (which determine the width of the Gaussian envelope along the x- and y-axes in the spatial domain). Given an input image $I(x, y)$, the filtered image $f(x, y)$ is the result of convolving $I(x, y)$ and $g(x, y)$. Tuning Gabor filters to specific frequencies and directions can enable them to detect both local orientation and frequency information from an image (Jones and Palmer, 1987; Weldon et al., 1996). We can assume that local image regions are spatially

Table 2.1: The mathematical expressions of GLCM features

No.	Feature name	Mathematical expression
1	Autocorrelation	$\sum_i \sum_j (i.j)p(i,j)$
2	Contrast	$\sum_i \sum_j \|i-j\|^2 p(i,j)$
3	Correlation 1	$\sum_i \sum_j \frac{(i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x \sigma_y}$
4	Correlation 2	$\sum_i \sum_j \frac{(i.j)p(i,j)-\mu_x \mu_y}{\sigma_x \sigma_y}$
5	Cluster prominence	$\sum_i \sum_j (i+j-\mu_x-\mu_y)^4 p(i,j)$
6	Cluster shade	$\sum_i \sum_j (i+j-\mu_x-\mu_y)^3 p(i,j)$
7	Dissimilarity	$\sum_i \sum_j \|i-j\|.p(i,j)$
8	Energy	$\sum_i \sum_j p(i,j)^2$
9	Entropy	$-\sum_i \sum_j p(i,j).log(p(i,j))$
10	Homogeneity 1	$\sum_i \sum_j \frac{p(i,j)}{1+\|i-j\|}$
11	Homogeneity 2	$\sum_i \sum_j \frac{p(i,j)}{1+\|i-j\|^2}$
12	Maximum probability	$\max_{i,j} p(i,j)$
13	Sum of squares	$\sum_i \sum_j (i-\mu)^2 p(i,j)$
14	Sum average	$\sum_{i=2}^{2N_g} i.p_{x+y}(i)$
15	Sum entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i).log(p_{x+y}(i))$
16	Sum variance	$\sum_{i=2}^{2N_g} (i-sum\ entropy)^2.p_{x+y}(i)$
17	Difference variance	$\sum_{i=0}^{N_g-1} i^2.p_{x-y}(i)$
18	Difference entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i).log(p_{x-y}(i))$
19	Information measure of correlation 1	$\frac{HXY-HXY1}{max(HX,HY)}$
20	Information measure of correlation 2	$\sqrt{(1-exp[-2(HXY2-HXY)])}$
21	Inverse difference normalized	$\sum_i \sum_j \frac{p(i,j)}{1+\|i-j\|^2/N_g}$
22	Inverse difference moment normalized	$\sum_i \sum_j \frac{p(i,j)}{1+(i-j)^2/N_g}$

Table 2.2: Expressions used to calculate GLCM features

Notation	Expressions	Explication
μ	Mean value of $p(i,j)$	The mean of the entire normalized GLCM
$p_x(i)$	$\sum_{j=1}^{N_g} p(i,j)$	i^{th} entry in the marginal-probability matrix obtained by summing the rows of $p(i,j)$
$p_y(i)$	$\sum_{i=1}^{N_g} p(i,j)$	i^{th} entry in the marginal-probability matrix obtained by summing the columns of $p(i,j)$
μ_x	$\sum_{i=1} \sum_{j=1} i.p(i,j)$	Mean of p_x
μ_y	$\sum_{i=1} \sum_{j=1} j.p(i,j)$	Mean of p_y
σ_x^2	$\sum_{i=1} \sum_{j=1} (i-\mu_x)^2.p(i,j)$	Variance of p_x
σ_y^2	$\sum_{i=1} \sum_{j=1} (j-\mu_y)^2.p(i,j)$	Variance of p_y
$p_{x+y}(k)$	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j),\ k=2,3,\ldots,2N_g$	Accumulate the probability matrix entries $p(i,j)$ that correspond to the sum of a set of pairs of gray levels
$p_{x-y}(k)$	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j),\ \|i-j\|=k,\ k=0,1,\ldots,N_g-1$	Accumulate the probability matrix entries $p(i,j)$ that correspond to the difference of a set of pairs of gray levels
HX	$-\sum_i p_x(i).log(p_x(i))$	Entropy of p_x
HY	$-\sum_i p_y(i).log(p_y(i))$	Entropy of p_y
HXY	$-\sum_i \sum_j p(i,j).log(p(i,j))$	Entropy of $p(i,j)$
$HXY1$	$-\sum_i \sum_j p(i,j).log(p_x(i)p_y(i))$	Expression similar to entropy equation used to compute information measure of correlation 1
$HXY2$	$-\sum_i \sum_j p_x(i)p_y(i).log(p_x(i)p_y(i))$	Expression similar to entropy equation used to compute information measure of correlation 2

homogeneous and use the mean and standard deviation of the magnitude of the filter responses to represent the region for classification. The feature vector f is constructed using the mean μ_{mn} and the standard deviation σ_{mn} of the filtered images, so $f = [\mu_{11}\ \sigma_{11}\ \mu_{12}\ \sigma_{12}\ \mu_{13}\ \sigma_{13} \ldots \mu_{mn}\ \sigma_{mn}]$. In this expression m and n are the number of scales and orientations of the filter, respectively.

2.3.3 Classification

Classification models are constructed using a labeled training set of the form $(x_i, y_i), i = 1, 2, \ldots k$, where $x_i \in \mathbb{R}^n$ are the feature values, $y_i \in \{1, -1\}$ is the binary classification, n is the number of features and k is the number of samples. The feature vectors are normalized in the interval $[0, 1]$, to prevent attributes with higher numeric ranges from dominating those with lower numeric ranges. The trained model can predict the class of new unlabeled instances.

2.3.3.1 k-nearest neighbor (k-NN)

The k-NN is a well-known classification method that is used in many applications (Altman, 1992). Given an instance, k-NN looks for the closest training points with respect to a particular distance metric (e.g. Euclidean, Manhattan, Minkowski and Hamming), and then it uses its labels to classify the instance by a majority vote.

2.3.3.2 Linear discriminant analysis (LDA)

The LDA classifier attempts to maximize the ratio of inter-class variance to the intra-class variance in any particular dataset thereby guaranteeing maximal separability (Scholkopft and Mullert, 1999). LDA considers maximizing the following objective:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{2.2}$$

In this equation, S_B is the "between classes scatter matrix" and S_W is the "within classes scatter matrix".

2.3.3.3 Support vector machine (SVM)

A SVM is a supervised learning classifier that discriminates between two classes by finding a hyperplane that separates them. Given a labeled training set of the form $(x_i, y_i), i = 1, 2, \ldots k$, where $x_i \in \mathbb{R}^n$ are the feature values, $y_i \in \{1, -1\}$ is the binary classification, n is the number of features and k is the number of samples, the SVM classifier solves the following optimization problem:

$$\begin{aligned} &\|\omega\|^2_{\omega,\xi} + C\sum_{i=1}^{k}\xi_i \\ &\qquad s.t. \quad y_i(\omega^T\phi(x_i) + b) \geq 1 - \xi_i, \\ &\qquad\qquad \xi_i \geq 0. \end{aligned} \tag{2.3}$$

In this expression, the soft margin parameter C tells the SVM optimization process how much is needed to avoid misclassifying each training instance. The weight vector ω is normal to the separating hyperplane. The parameter ξ is used to give a degree of flexibility to the algorithm when fitting the data and b represents the bias.
During the optimization process of the SVM the training data x_i are mapped into a higher dimensional space using a kernel function, $K(x_i, x_j) = (\phi^T(x_i).\phi(x_j))$. SVM uses the kernel trick, by which the data become linearly separable in the new space. The SVM classifier finds the hyperplane with a maximum margin of separation between the classes in the new higher dimensional space. In the case of a *linear SVM (LSVM)* classifier, ϕ refers to a dot product. In a *non-linear SVM (NLSVM)* the classifier function is formed by non-linearly projecting the training data of the input space to a feature space of a higher dimension by using a kernel function. The radial basis function (RBF) is widely used as a mapping kernel. The RBF can be defined as follows:

$$K(x_i, x_j) = exp\left(-\gamma\|x_i - x_j\|_2^2\right) \tag{2.4}$$

In this expression $\gamma = 1/2\sigma^2$, $\|x_i - x_j\|_2^2$ is the squared Euclidean distance between the two feature vectors x_i and x_j, and σ is a free parameter.

2.3.3.4 Random forests (RF)

RF is an ensemble learning method that operates by constructing a multitude of decision trees at the training stage and producing the class that is the mode of the classes output by the individual trees in the test stage (Breiman, 2001). Decision trees are a supervised learning method used for classification and regression. The aim of this method is to construct a model that predicts the value of a target by learning simple decision rules derived from the data. An RF can be built by randomly sampling a training data subset for each decision tree (as in Bagging (Breiman, 1996)). According to (Breiman, 2001), each tree is trained using 2/3 of the total training data. Then, each tree is validated using the samples that have not been used to build that tree (*out-of-bag error* estimation).

2.3.3.5 Multi-layer perceptron (MLP)

An MLP is a network of simple neurons called perceptrons. Each perceptron determines a single output from multiple numerical inputs by setting a weight for each input and then combining them. The combined value is then sent to a nonlinear activation function (Hornik et al., 1989). This operation can be formulated using the following expression:

$$y = \phi(\sum_{i=1}^{n} w_i x_i + b) \tag{2.5}$$

In this expression, y is the output of the perceptron, x is the vector of inputs, w is the vector of weights, b is the bias and ϕ is the activation function. The common activation functions are the logistic sigmoid $\phi(z) = 1/(1 + e^{-z})$ and the hyperbolic tangent $\phi(z) = \tanh(z)$. Indeed, a single perceptron is not very useful because of its limited mapping ability. A typical MLP network consists of an input layer, one or more hidden layers, and an output layer. The input propagates through the network layer-by-layer to produce the final output. MLP networks are widely used in supervised learning problems, which can be solved with the back-propagation algorithm. This algorithm consists of forward and backward passes. In the forward pass, the predicted outputs corresponding to the given inputs are evaluated. In the

backward pass, partial derivatives of the cost function with respect to the different parameters are propagated back through the network.

2.3.4 Evaluation

In medical diagnosis tests, sensitivity (also known as recall) and specificity are usually calculated. Sensitivity measures the proportion of actual positive instances which are correctly identified. Specificity measures the proportion of negative instances which are correctly identified. A perfect breast cancer CAD would have a sensitivity and specificity of 100%. In addition, precision, accuracy and F-score are used to evaluate CAD systems. The following terms are defined to calculate the aforementioned metrics:

- True Positive (TP): positive instance classified as positive.
- True Negative (TN): negative instance classified as negative.
- False Positive (FP): negative instance classified as positive.
- False Negative (FN): positive instance classified as negative.

Table 2.3 presents the mathematical expressions of sensitivity, specificity, precision, accuracy and F-score. In the literature, the performance of CAD systems is commonly measured in terms of the area under the curve (AUC) of the receiver operating characteristic (ROC). The ROC analysis is used to avoid selecting a single threshold for classification. All the possible thresholds are scanned and the effect on the true positive rate $TP/(TP + FN)$ (sensitivity) and the false positive rate $FP/(FP+TN)$ (1-specificity) is recorded. In this way, the ROC curve describes the performance of a model throughout the range of classification thresholds (Fawcett, 2006). Fig. 2.6 shows an example of ROC with an AUC of 0.79.

Table 2.3: The mathematical expressions of sensitivity, specificity, precision, accuracy and F-score

Metric	**Expression**
Sensitivity	$TP/(TP + FN)$
Specificity	$TN/(TN + FP)$
Precision	$TP/(TP + FP)$
Accuracy	$(TP + TN)/(TP + TN + FP + FN)$
F-score	$2TP/(2TP + FP + FN)$

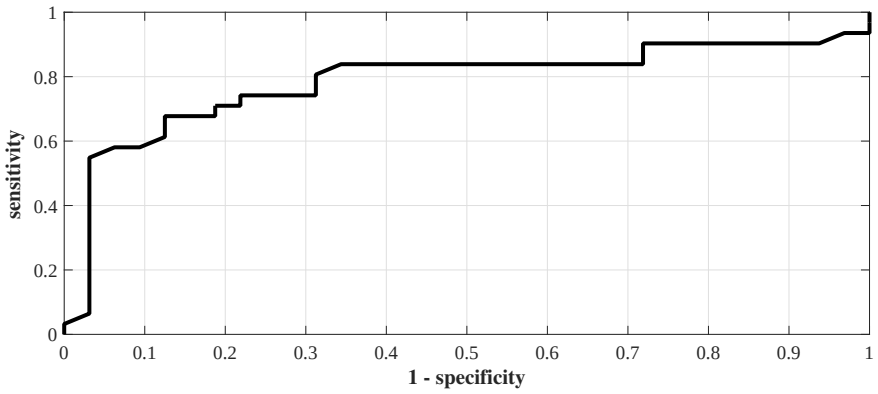


Figure 2.6: ROC curve

2.3.4.1 Cross validation

Cross validation has been widely used to evaluate machine learning algorithms. It is used to give an indication of how well the model will predict unseen data. Cross validation is done by partitioning a dataset and using a subset to train the algorithm and the remaining data to test it. The common cross validation techniques are:

- **k-fold cross validation.** In this technique, the data are randomly sorted and divided into k folds (a common value of k is 10). One of the folds is used for testing and the remaining folds for training the algorithm. This scenario is repeated k times.
- **Leave-one-out-cross validation (LOOCV).** In this technique, the data is partitioned using the k-fold approach where k is equal to the total number of observations in the data.
- **Holdout.** This technique divides the data into exactly two subsets of specified ratio for training and testing.

2.3.4.2 Statistical analysis

Statistical analysis is commonly used to measure the statistical significance of the results of CAD systems. A result is statistically significant when a p-value is less than the significance level (α). The p-value is the probability of obtaining at least as extreme results given that the null hypothesis is true. The significance level α is the probability of rejecting the null hypothesis given that it is true (Lowry, 2014).

Statistical methods can be parametric or nonparametric. Parametric methods have a number of parameters; whereas, nonparametric methods are statistics not based on parameterized families of probability distributions (i.e., parameters are determined by the training data). Several parametric statistical methods have been used to evaluate CAD systems, such as Student's t-test and ANOVA. Indeed, Student's t-test can be applied when the data follows a normal distribution. In turn, many studies have used nonparametric methods to determine the statistical significance of the results of CAD systems. The Wilcoxon signed-rank test, which is an example of these methods, can be used as an alternative to the t-test when the population of the data does not follow a normal distribution.

Furthermore, the *kappa* statistic has been used to evaluate the results of breast cancer CAD systems. This statistic compares the accuracy of the system to the accuracy of a random system. It is used to assess the inter-rater reliability when observing categorical variables (κ= 1 if there is *full agreement*; κ= 0 if there is no agreement). The kappa coefficient (Landis and Koch, 1977) is defined as

$$\kappa = \frac{P_a - P_e}{1 - P_e} \tag{2.6}$$

In this equation, P_a is the relative observed agreement among raters and P_e is the hypothetical probability of the agreement derived by chance using the observed data to calculate the probabilities of each observer randomly saying each category.

2.4 Analysis of the temporal evolution of breast tumors

The evolution of breast tumors can be studied through motion analysis methods. In this thesis, we study image registration and optical flow methods.

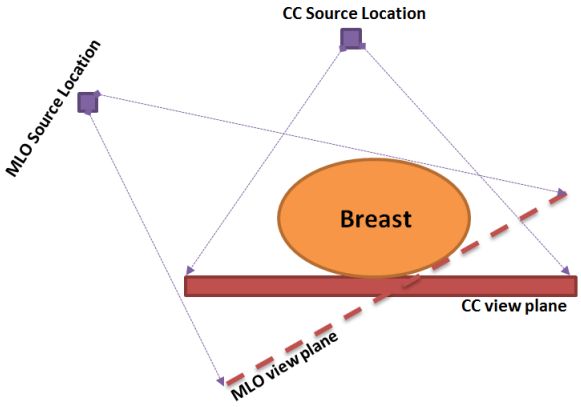


Figure 2.7: Acquisition of CC and MLO views

2.4.1 Mammogram registration

A lesion may not be recognizable from one view, and the appearance may be different on MLO and CC views (see Fig. 2.7). Registration helps doctors analyze and visualize mammograms. The comparison of mammograms requires a registration (alignment) between the mammogram images. *The registration process* aligns the coordinates of the current mammogram (*template*) to the coordinates of a previous one (*reference* mammogram). Two main categories of image registration methods can be identified: *intensity-based* and *feature-based*. Intensity-based registration depends on the image pixel values. The change in illumination between mammograms and the mapping model affects the accuracy of these methods. In turn, feature-based methods depend on the features extracted from the mammograms. Control points are the most commonly used features in the registration of mammogram images. Actually, it is difficult to extract consistent features from the mammograms because the appearance of the breast region depends heavily on the compression of the breast.

The registration process has three main components: a similarity function, deformation model and optimizer. The choice of each component depends on some assumptions related to the applications. The *similarity measure* determines the quality of the alignment between the images. These measures are divided into intra-modality (registering images from the same modality) and inter-modality (registering images from different modalities) measures. The sum of square

differences (SSD) and the cross correlation (CC) are commonly used in intra-modality registration methods. In turn, mutual information (MI) and normalized mutual information (NMI) are used in inter-modality registration methods. The large deformations between mammograms are the main problem of any mammogram registration method. Deformations mainly occur due to change in breast compression, acquisition time, position, age and the imaging view angle. The registration process should compensate the deformations between the reference and template mammograms and correctly align the interior structures in breast regions. The *deformation models* can be rigid or elastic. Rigid deformation models, such as affine transformation, are applied to the whole image and are used to compensate the global motion in the breast region. In turn, elastic models, such as the elastic body spline model (EBS), the B-spline-based model and thin plate spline model (TPS) are used to compensate for the local motion in the breast region and they may be applied to selected parts of the image (Crum et al., 2014).

An *optimization* process is essential if the values of the parameters of the transformation model used that maximize the similarity measure are to be found. The simplest optimization technique that guarantees a global optimal solution is the exhaustive search. The computational complexity of the exhaustive search is high, so its usage is limited to an optimization problem with a small search space (e.g. registration methods based on the translational model). The common optimization methods used in the literature are gradient descent, genetic algorithm, Newton-Raphson, simulated annealing and Levenberg-Marquardt optimization (Brown, 1992).

2.4.2 Optical flow

In the last three decades, several approaches have been used to calculate the optical flow. In general, an optical flow algorithm characterizes the spatial arrangement of the objects and the rate of change of this arrangement in two successive images of the same visual scene. An optical flow algorithm produces the flow velocity vector $w =: (u, v)^\intercal$, where u is the displacement in the horizontal direction and v is the

displacement in the vertical direction. Optical flow can be applied locally as proposed in (Lucas and Kanade, 1981) or globally as proposed in (Horn and Schunck, 1981), although most optical flow techniques are based on the latter approach.

Assume I1 and I2 are two images captured at times t and $t+1$, respectively, $x := (x, y)^\intercal$ is a point in the image domain, and $w := (u, v)^\intercal$ includes the displacement vectors. Several optical flow methods are based on brightness constancy and small motion assumptions. Brightness constancy can be formulated with the expression

$$I(x, y, t) \approx I(x + u, y + v, t + dt) \tag{2.7}$$

where $I(x, y, t)$ is the intensity value of a pixel located at position (x, y) at time t. Assuming small motion, the corresponding point of a pixel can be found within a very small neighborhood.

In the case of follow-up mammograms, the variations in breast compression and radiation dose (which produces changes in illumination) between successive mammograms violate both the brightness and the small motion constancy assumptions. To cope with the large deformations that appear in mammograms (due to variation in compression), large displacement optical flow methods are required. A coarse-to-fine image warping was introduced to overcome large displacements (a motion larger than one pixel). This can be used jointly with many robust objective functions that have been proposed to reduce the effect of illumination change (Sun et al., 2010).

Coarse-to-fine technique (CTF). For each image, an image pyramid is first built (see Fig. 2.8). An image pyramid involves a low pass filtering (e.g., a Gaussian filter) and down-sampling the image. Starting from the lowest resolution level of the pyramid (coarse level), the optical flow is calculated. This calculated optical flow is then passed to the next highest resolution level as an initial estimate. The optical flow estimated at a coarse level is used to warp the second image towards the first one at the next fine level, and a flow increment is calculated between the first image and the warped second image. This process is repeated until the coarse level is reached.

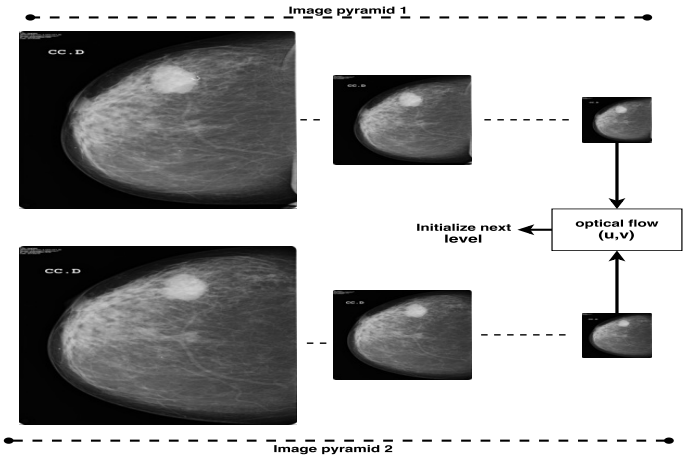


Figure 2.8: Coarse-to-fine technique (in this example, each image pyramid consists of three levels)

2.5 Breast cancer databases

There is a limited number of publicly available breast cancer datasets. In this subsection, we summarize the ones used in this thesis.

- **Mini-MIAS (Suckling et al., 1994).** A film-screen (digitized) database that contains 322 images (pgm format) of 161 women in MLO view only. The mini-MIAS database was created from the original MIAS database (digitised at 50 μ pixel edge) by down-sampling it to 200 μ pixel and clipping/padding it to a fixed size of 1024×1024 pixels. The down-sampling was done by the authors of the database. A GT was prepared by experienced radiologists and confirmed using a biopsy test. The GT of the mini-MIAS mammograms shows the location of the abnormality, the radius of the circle which contains the abnormal region, the characteristics of the background tissues, the breast density of each image (fatty, glandular or dense) and the severity of each abnormality. The dataset is available at http://peipa.essex.ac.uk/info/mias.html.
- **INbreast (Moreira et al., 2012).** A full-field digital mammographic (FFDM) database that contains images of 115 women. A total of 410 images were collected between April 2008 and July 2010. The images were acquired using MammoNovation Siemens FFDM equipment with a solid-state detector (pixel size of 70μ, 14-bit contrast resolution). The sizes of the acquired

images are 3328×4084 or 2560×3328 pixels. The image size depends on the compression plates used in the acquisition process (the amount of compression varied according to the breast size of each woman). The images collected were saved in the digital imaging and communication in medicine (DICOM) format. The GT was prepared by an experienced radiologist and validated by a different one. The annotations were generated using OsiriX (an open-source picture archiving application) and a communication system (PACS) workstation.

- **Breast ultrasound sequences database.** A database that contains 31 malignant and 28 benign BUS videos. It is part of a clinical database of ultrasonic radio frequency strain imaging data that was created by the Engineering Department, Cambridge University. It is available at http://mi.eng.cam.ac.uk/research/projects/elasprj/.
- **Database of breast thermograms (Silva et al., 2014).** A database that contains 148 thermograms of 148 women, both healthy and sick. We collected the dataset from the Proeng database, which is available at http://visual.ic.uff.br/en/proeng/. In a thermogram each pixel corresponds to the temperature of the acquired scene. The range of temperatures is associated with the gray-scale of the images. The size of the images is 640×480 pixels. The thermograms were captured by a FLIR thermal camera, model $SC620$, which has a sensitivity of less than 0.04^o and captures temperatures from -40^oC to 500^oC.

2.6 Conclusion

In this chapter, we have presented a background on breast cancer screening methods, ROI segmentation, texture analysis methods, classification methods, CAD evaluation, image registration, optical flow and breast cancer datasets. In the next section we present several methods for analyzing breast cancer in mammograms.

Part II

Analysis of breast cancer in mammograms

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

CHAPTER 3

Mass analysis in mammograms using texture methods

3.1 Introduction

Several methods have been proposed for automatic breast mass detection using computer vision and machine learning techniques. These methods can be classified as *supervised* or *unsupervised*. In the supervised methods, which are more accurate, feature vectors are extracted from the ROIs. The extracted features can be related to *texture*, *statistic*, *position* and *geometry*. The normal and mass regions are the most interesting ROIs in breast mammograms. The extracted features are used to train a model, which is used to predict the class of unknown instances. Unfortunately, supervised methods usually present a high number of false positive detections.

In practice many factors influence the false positive mass detection rates in CAD systems, and these should be carefully analyzed to reduce the percentage of errors. In previous studies we have identified the following factors:

- The ability of given feature extraction methods to distinguish between mass and normal ROIs.
- The robustness of the classifier used for training and prediction.
- Breast density: dense breast tissue usually produces light gray levels in mammograms, making them harder to interpret, specially in younger women. In practice, it is difficult to detect masses in mammograms with high densities.
- Quality of the mammograms: both noise and physical artifacts may degrade the overall performance of the mass detection algorithms.

Although several feature extraction methods have been proposed for mammogram image analysis, improving the classification results remains a challenge and an open problem. In this chapter, we analyze the performance of various texture analysis methods for breast mass detection in an attempt to reduce the *false positives* and, therefore, the number of *unnecessary biopsies*. We used local binary patterns, histogram of oriented gradients, co-occurrence matrix features, local grey level appearance and Gabor filters. The novelty of this study is that we propose local directional number patterns (LDN) as a new texture feature extraction method for breast mass detection. Once the features have been extracted by each method, a binary classifier (mass/normal tissue) is trained. As mentioned above, these models have shown robust capabilities in classification problems and they have been widely used in medical applications. We have also analyzed possible combinations of different texture analysis methods (both considering the majority output of the individual classifiers and building new models on the concatenation of features provided by different methods).

3.2 Related studies

The literature describes *single feature* and *multiple feature* breast cancer CAD systems. Several studies have proposed the use of a *single feature family*. Oliver et al. (2007) used the histogram of the Local Binary Pattern (LBP) to reduce the number of false positives in breast mass detection. They used an SVM for classification. Unfortunately, LBP may assign the same pattern to a pixel in a tumor region and another pixel in normal dense tissue. This leads to a noticeable percentage of false detections. The HOG has also been used for breast mass detection (Pomponiu et al., 2014). The HOG descriptor can be used to extract features with which to train an SVM classifier. The cell size and the number of cells per block need to be optimized because if the block size is unsuitable, the same HOG descriptor may be produced for a dense normal block and a tumor block leading to high false detections. Qian et al. (1999) used a multi-resolution and multi-orientation wavelet transform for mass detection and spiculation analysis in breast mammograms. They observed that the traditional wavelet can not extract the directional information that characterizes the spiculations.

Some breast cancer CAD systems used *multiple feature families*. In (Christoyianni et al., 2002), grey levels, texture and features related to independent component analysis are used to train a Neural Network classifier. The usage of grey level creates a dilemma in the classification stage as the dense normal pixels and the tumorous pixels have similar intensity values (i.e. they have the same visual characteristics). A comparison between Haralick's features, wavelet-based features and multi-wavelet-based features was presented in (Soltanian-Zadeh et al., 2004). It was concluded that the multi-wavelet features followed by shape features yielded the best ROC results, but they still produce a noticeable number of false positives. The main reason for this is that Haralick's features depend on the co-occurrence matrix, which counts the number of pixels that have the same value at a certain distance and angle. Unfortunately, the same co-occurrence matrix can be achieved for a tumor ROI and a normal ROI in a dense mammogram. Moreover, a mass

detection approach used a Growing Neural Gas algorithm to perform a segmentation step (de Oliveira Martins et al., 2009). A set of shape measures are computed for each segmented region to suppress bad mass candidates and texture measures are obtained from Ripley's K function. The shape measures cannot characterize the mass regions correctly in dense mammograms. Moreover, the segmentation step may merge a region that contains dense normal tissue with a region that contains tumorous tissue, thus yielding a high number of false positives.

The classification stage in breast cancer CAD systems can be *unsupervised* or *supervised*. Zheng (2010) proposed an *unsupervised* breast mass detection method that used a circular Gaussian filter followed by a threshold to segment the input mammogram. Afterwards, a Gabor filter is used to extract features from these segments. There are many false positives due to the use of a threshold in the segmentation step, especially in dense mammograms in which different regions can be merged. Bellotti1 et al. (2006) proposed an automatic breast mass detection approach based on a *supervised* classification method. They used an edge-based segmentation algorithm to separate suspicious regions. In addition, second order measures obtained from the co-occurrence matrix were used to describe the texture of each segment. Finally, an Artificial Neural Network is used for classification. The descriptors used produced a sensitivity of 80%. This result indicates that the co-occurrence matrix features yield a high number of false positives, and they do not accurately describe breast tissues. A feature selection procedure based on a genetic algorithm is used to select the suitable set of features, which are fed into a k-NN classifier. On the other hand, de Oliveira Martins et al. (2009) used both supervised and unsupervised learning in breast mass detection. They used the K-means algorithm for mammogram segmentation, and the co-occurrence matrix features to describe the texture of the segmented structures. Finally, they classified these segmented structures using an SVM classifier. The main drawback of this method is that the co-occurrence matrix can produce the same descriptor for a normal cluster in a dense region and a tumorous cluster, yielding low classification accuracy. Indeed, the literature has been unable to define an optimal set of features for the task

of mass/normal breast tissue classification, which means that the methods proposed in the literature do not accurately describe breast tissues, especially when dense breasts are considered.

3.3 Methods

Breast masses are usually brighter than normal tissues in mammograms, and they are defined by the characteristics of their shapes and margins. Mass detection is quite difficult in dense mammograms, in which normal tissues are also bright and they cover the masses. If suitable descriptors for characterizing breast tissues are to be selected, the most significant features of breast masses should be considered.
In this study, we have considered the texture analysis methods which have most often been applied in the field: for example, LBP, HOG, GLCM features, Gabor filters, and local grey level appearance (LGA). We also introduce LDN as a new analysis method in this area and combine different features in an attempt to improve the performance of breast mass detection. Indeed, our analysis is twofold. First we analyse each texture method individually, considering two ROI ratios, with classification methods such as k-NN, LDA, LSVM, NLSVM and RF. And second, we analyze possible combinations of different texture analysis methods (considering the majority output of the individual classifiers and building new models on the concatenation of features provided by different methods).

3.3.1 Feature extraction

This subsection explains the texture analysis methods used and the selection of the parameters of each method. In section 2.3.2, we explained the local binary pattern (LBP), the histogram of oriented gradients (HOG), the grey level co-occurrence matrix (GLCM) features and Gabor filters.

Local directional number pattern (LDN). LDN was proposed in (Ramirez Rivera et al., 2013) for face analysis. They showed that LDN is better than LBP because it can detect changes in regions producing different 6-bit

codes, while LBP can produce the same pattern for pixels in different regions. Taking into account the advantages of LDN over LBP and the previous use of LBP in mass detection (Oliver et al., 2007), we felt that LDN could be a good method for this problem. In LDN, the edge responses are computed in eight different directions by convolving the Kirsch compass masks with the ROI (see Fig. 3.1). The authors of LDN chose the location of the top positive and negative edge responses to generate a meaningful descriptor for each pixel (6-bit code). An example of LDN is presented in Fig. 3.1. The location codes that return the maximum and minimum responses are concatenated to form the LDN code of the marked pixel. LDN encodes the directions in a binary code shorter than LBP's, and it has been shown to perform well in such fields as face expression analysis. The feature vector of LDN is calculated by dividing each ROI into four sub-regions, and the final LDN feature vector ($4 \times 64 = 256$ dimensions) is formed by concatenating the histograms of the four sub-regions. This step increases the description ability of LDN, as it adds global spatial information to the final descriptor.

Local binary pattern (LBP). Many variants of LBP have been proposed in the literature: for example, robust local binary pattern (RLBP), fuzzy local binary pattern (FLBP) and center symmetric local binary pattern (CSLBP). We computed the histogram with the frequency of uniform LBPs for each ROI. We used 3×3 local windows to generate the uniform LBP for each ROI. Each ROI was divided into four regions, and the final LBP histogram was created by concatenating the histograms of the four regions. The final dimension of the LBP feature vector was 236 (4×59).

Robust local binary pattern (RLBP) divides each 8-bit LBP binary code into sets of three overlapped bits (Chen et al., 2013). Then, if '010' or '101' are found, they are replaced by '000' or '111' respectively. The main problem of Chen's method is that it converts a natural non-uniform pattern to a uniform pattern, thus distorting the overall description, because the wrong correction will increment the number of uniform patterns in the final LBP histogram. The feature vector of RLBP is calculated by dividing each ROI into four sub-regions, and then the

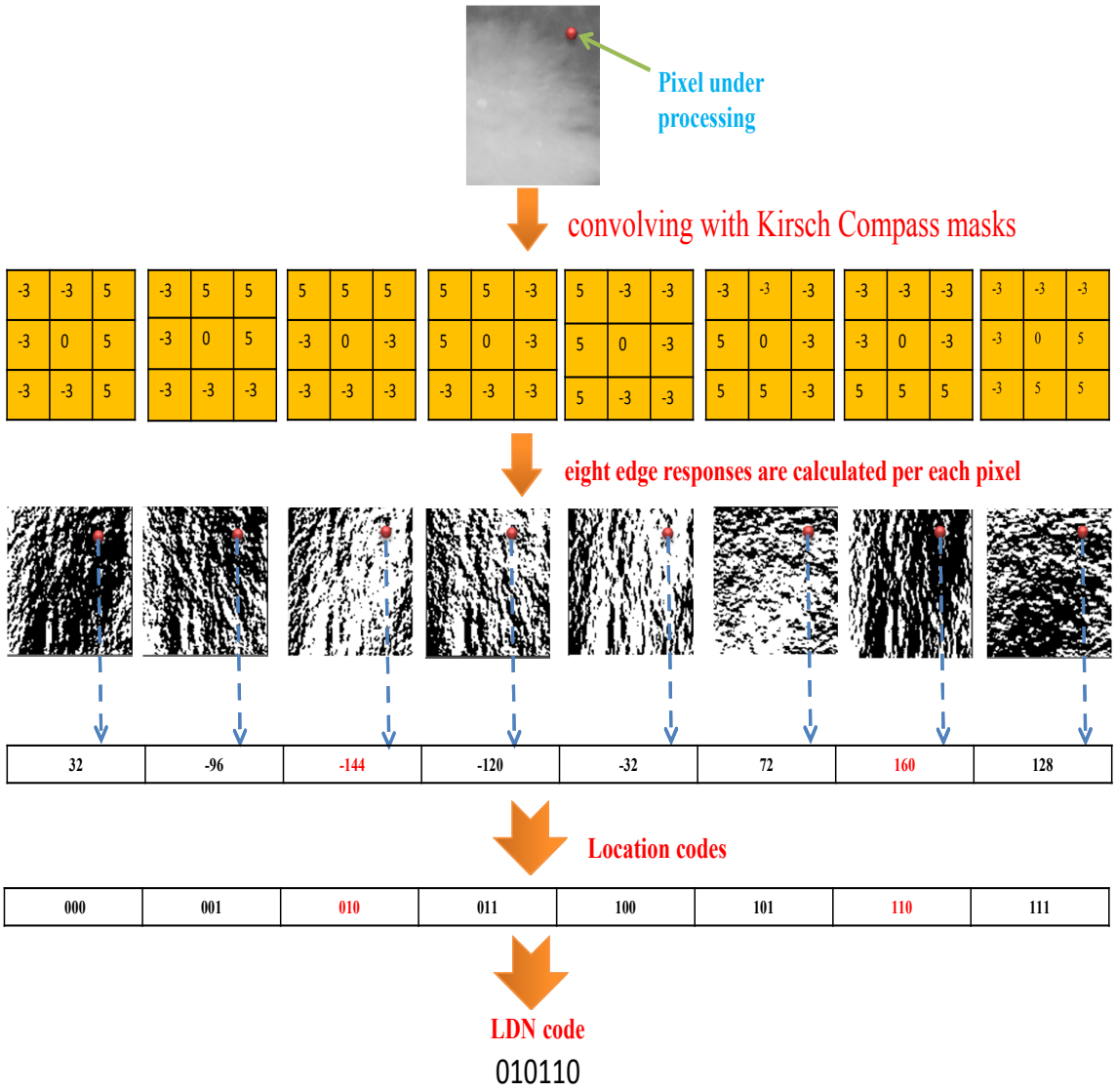


Figure 3.1: Example of LDN binary code generation

resulting histograms are concatenated. The final dimension of the RLBP feature vector was 236 (4×59).

Centre symmetric local binary pattern (CSLBP) is an extension of the original LBP operator (Heikkilä et al., 2009). CSLBP compares the center symmetric pairs of pixels as illustrated in Fig. 3.2. It describes each pixel in the ROI using 4-bits. The feature vector of CSLBP is calculated by dividing each ROI into four sub-regions, and then the resulting histograms are concatenated (the dimension of the CSLBP feature vector was 64 (4×16)).

Fuzzy local binary pattern (FLBP) is an extension of LBP proposed in (Iakovidis

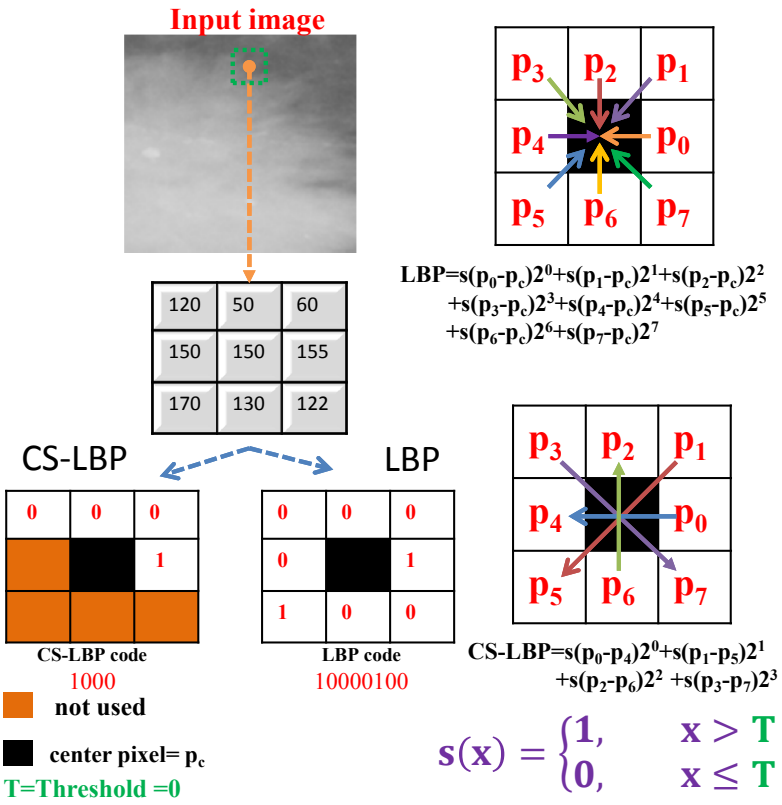


Figure 3.2: Example of LBP and CS-LBP code generation

et al., 2008). FLBP incorporates fuzzy logic in the representation of local patterns. Assume that d_i is the difference between a given pixel p_i in a local neighborhood and its centre p_c. A membership function is used to determine the degree of d_i to be '1' or '0' (instead of comparing d_i with zero as in the case of LBP). Fuzzification allows each local neighborhood to contribute to more than a single bin in the distribution of the LBP. In (Iakovidis et al., 2008), FLBP is evaluated by classifying ROIs extracted from thyroid ultrasound images as nodules or non-nodules. In our experiments, the tuned value of the fuzzy parameter was 15, and the histogram of FLBP is used as a feature vector (256 dimensions).

HOG features. We used a 3×3 cell size, 8×8 cells for the block size, and a 9-bit histogram. The dimension of the HOG feature vector was 576 ($9 \times 8 \times 8$).

GLCM features. To increase the description given by GLCM, spatial information should be added, so we divided each ROI into four sub-regions. For each sub-region we calculated 12 GLCMs by combining four orientations (0^o, 45^o, 90^o and 135^o) and three distances (3, 5 and 10 pixels). From each GLCM we calculated 22 texture

features, and we concatenated all the features calculated for each GLCM for all the sub-regions into one feature vector ($4 \times 4 \times 3 \times 22 = 1056$ dimensions).

Gabor filters. Gabor feature vector f is constructed using the mean μ_{mn} and the standard deviation σ_{mn} of the filtered images, so $f = [\mu_{11}\ \sigma_{11}\ \mu_{12}\ \sigma_{12}\ \mu_{13}\ \sigma_{13} \ldots \mu_{mn}\ \sigma_{mn}]$. In this expression m and n are the number of scales and orientations of the filter, respectively. We used 4 scales and 6 orientations to calculate the responses of Gabor filters ($4 \times 6 = 24$ responses). The feature vector is constructed by concatenating all the resulting means and standard deviations ($4 \times 6 \times 2 = 48$ dimensions).

Local grey level appearance (LGA). The LGA approach was presented in (Zwiggelaar, 2010). In order to calculate a N_b-bin LGA histogram, the input image grey level range is reduced to N_b, and then a $w \times w$ local neighborhood is extracted at each pixel in the input image. A unique LGA number corresponding to each local neighborhood is computed as follows:

$$LGA(i,j) = \sum_{i,j} N_b^{s(i,j)} I_d(i,j) \tag{3.1}$$

where N_b is the number of bins in the LGA histogram, $s(i,j) = 0, 1, 2 \ldots N-1$ is the sequence number of the pixels, N is the number of pixels in the local neighborhood, and $I_d(i,j)$ is the reduced resolution grey level value at position (i,j). We used a 3×3 local window (N=9). The LGA histogram can be built by counting the occurrence of LGA numbers. We set the number N_b of bins in the LGA histogram to 8. The feature vector of LGA is calculated by dividing each ROI into four sub-regions, and the final LGA feature vector is formed by concatenating the histograms of the four sub regions ($4 \times 8 = 32$ dimensions).

3.3.2 Classification

In this study, we used the following classifiers to classify the ROIs into normal or mass:

- **k-NN classifier.** We used the Euclidean distance to determine the nearest

neighbours of the test instance (it was empirically determined that the optimum k was 3).

- **LDA classifier.** We used its MATLAB-based implementation (MATLAB R2013a). In our experiments, we found that the best discriminant type for our problem is the *pseudoLinear* one.
- **SVM classifier.** In this thesis, the SVM classifier is implemented with Matlab and the libSVM library (Chang and Lin, 2011). In addition, a grid search algorithm was performed to find the optimal values of the kernel's parameter γ and the soft margin parameter C. In the grid search algorithm, a k-fold cross validation was used to divide the training data into k sets. Then *k-1* sets were used to train the classifier and the remaining set was used to test the trained model. Many pairs of (C,γ) were tried and the one with the highest cross-validation F_1 score (the harmonic mean of precision and recall) was selected. As recommended in (Hsu et al., 2003), the grid search algorithm tried many exponentially growing sequences of C and γ. Possible values for C were $(2^{-5}, 2^{-3}, \ldots, 2^{15})$ and for γ were $(2^{-15}, 2^{-13}, \ldots, 2^{3})$. After determining the best values for (C, γ), the SVM accurately classified the unknown instances. For further information about the details of the grid search, the libSVM guide may provide help (Hsu et al., 2003).
- **RF classifier.** We used a MATLAB based RF implementation. The optimum number of trees that stabilizes the *out-of-bag* error was 30.

3.4 Experimental results and discussion

Our analysis is twofold. First we analyze each texture method individually, considering two ROI ratios (the ratio between the number of mass ROIs and normal ROIs). We then present the result of combining these methods. We used two combination techniques: in the first we concatenated the features obtained by different methods, and trained a new classifier on the extended set of features, whereas in the second we used the individual classifiers obtained by different methods

and considered their majority opinion.

We used the mini-MIAS database in our experiments (Suckling et al., 1994). The GT of mini-MIAS shows the location of the abnormality, the radius of the circle enclosing the abnormal region and the characteristics of the background tissues. To generate ROIs we followed the procedure given in (García-Manso et al., 2013). In turn, the ROIs of normal tissue were selected randomly from the normal mammograms. Normal-tissue ROIs were created with sizes randomly ranging from the smallest to the largest found in the databases. In the case of the mini-MIAS database, 109 mass ROIs from the mass mammograms and 203 normal ROIs from the normal mammograms were generated. The ROIs generated are of different sizes because the tumors in the mini-MIAS images are also of different sizes, so all ROIs were resized into a common size (150×150 pixels). We used the k-fold technique to generate the training and testing datasets (in this study k is set 10). To evaluate the performance of each texture extraction method, we calculated the mean sensitivity and specificity across k-folds. A high sensitivity implies a small number of false positives so only a few of the biopsies performed would be unnecessary.

3.4.1 Analysis of the texture methods

Tables 3.1 to 3.5 show the sensitivity and specificity of each method over all the iterations of the 10-fold procedure considering the KNN, LDA, RF, LSVM and NLSVM classifiers, respectively[1]. In the following tables the methods which have sensitivities and specificities over 80% are highlighted. Table 3.1 indicates that LDN, LBP and HOG produce a good description of breast tissues as they have good sensitivity and specificity with the KNN classifier (the simplest classifier). Results are good with a balanced dataset, but sensitivity decreases when the dataset is imbalanced. Table 3.2 shows that the use of Gabor filters is the only method that produces good sensitivity and specificity with the LDA classifier for both balanced and imbalanced datasets. These results indicate that the LDA classifier is unsuitable

[1]In each cell of the tables, we present (*mean* $\pm$ *standard deviation*) over all the iterations of the 10-fold procedure.

for separating the feature space of other methods into mass and normal classes.

Table 3.1: Sensitivity and specificity of breast mass detection with LDN, LBP, RLBP, CSLBP, FLBP, HOG, CM, Gabor and LGA using KNN classifier (k=3)

Method	1/1 ROI		1/2ROI	
	Sensitivity	Specificity	Sensitivity	Specificity
LDN	**84.43**±0.0727	**90.33**±0.0540	79.74±0.0855	**93.59**±0.0191
LBP	**82.58**±0.0712	**97.14**±0.0206	**83.21**±0.0759	**97.48**±0.0159
RLBP	77.45±0.0550	91.28±0.0494	76.56±0.0691	98.07±0.0141
CSLBP	76.46±0.0870	93.28±0.0378	66.83±0.0548	92.27±0.0164
FLBP	72.17±0.0493	89.99±0.0409	53.17±0.1263	93.30±0.0509
HOG	**81.80**±0.0214	**92.45**±0.0473	46.60±0.1015	93.93±0.0165
GLCM	47.06±0.1029	74.23±0.0597	51.42±0.0613	77.78±0.0344
Gabor	74.84±0.1174	67.91±0.0502	54.08±0.0491	89.46±0.0305
LGA	58.78±0.1025	73.85±0.0285	45.41±0.0941	85.63±0.0239

Table 3.2: Sensitivity and specificity of breast mass detection with LDN, LBP, RLBP, CSLBP, FLBP, HOG, GLCM, Gabor and LGA using LDA Classifier

Method	1/1 ROI		1/2ROI	
	Sensitivity	Specificity	Sensitivity	Specificity
LDN	69.00±0.1969	80.00±0.0667	63.00±0.1337	69.50±0.0798
LBP	75.00±0.0850	80.00±0.1155	63.00±0.1636	75.00±0.0882
RLBP	73.00±0.1337	86.00±0.1265	74.00±0.1174	80.00±0.0850
CSLBP	77.00±0.1059	88.00±0.1135	74.00±0.1075	95.00±0.0408
FLBP	67.00±0.1160	63.00±0.2359	69.00±0.1792	66.00±0.1150
HOG	66.00±0.1647	76.00±0.1506	63.00±0.1829	87.50±0.0717
GLCM	74.00±0.1174	77.00±0.1494	75.00±0.1179	92.00±0.0587
Gabor	**83.00**±0.1337	**90.00**±0.0816	72.00±0.1135	**95.00**±0.0333
LGA	39.00±0.0876	86.00±0.0843	37.00±0.1703	94.50±0.0798

Table 3.3 indicates that the RF classifier gives good results with LDN, LBP, RLBP, CSLBP, FLBP and HOG with both balanced and imbalanced datasets. Furthermore, Table 3.4 indicates that LDN features are the most linearly separable ones as they produce the best sensitivity and specificity with the LSVM classifier with the balanced dataset, and a specificity of 99.0% with the imbalanced dataset. In turn, Table 3.5 shows that LDN, LBP, RLBP, and HOG have good sensitivity and specificity for both the balanced and imbalanced datasets with the NLSVM classifier.

On the basis of the previous experiments, we conclude that LDN yields the best sensitivities and specificities with the NLSVM and RF classifiers. LBP, RLBP and HOG also gave good results with the NLSVM and RF classifiers, but they have some

Table 3.3: Sensitivity and specificity of breast mass detection with LDN, LBP, RLBP, CSLBP, FLBP, HOG, GLCM, Gabor and LGA using the RF classifier (30 trees)

Method	1/1 ROI		1/2ROI	
	Sensitivity	Specificity	Sensitivity	Specificity
LDN	**81.00**±0.1370	**95.00**±0.0707	**80.00**±0.1764	**98.50**±0.0337
LBP	**87.00**±0.1494	**95.00**±0.0707	**84.00**±0.0516	**97.00**±0.0258
RLBP	**90.00**±0.1054	**94.00**±0.0699	**84.00**±0.0843	**96.50**±0.0337
CSLBP	**85.00**±0.0707	**95.00**±0.0707	**84.00**±0.1174	**98.00**±0.0258
FLBP	**84.00**±0.1075	**90.00**±0.1054	**84.00**±0.1430	**94.00**±0.0516
HOG	**83.00**±0.1160	**90.00**±0.1155	71.00±0.1524	**95.50**±0.0550
GLCM	73.00±0.1337	81.00±0.1449	70.00±0.1764	95.00±0.0408
Gabor	73.00±0.0949	86.00±0.0966	70.00±0.1700	94.50±0.0550
LGA	63.00±0.1767	80.00±0.1247	45.00±0.0972	93.00±0.0587

Table 3.4: Sensitivity and Specificity of breast mass detection with LDN, LBP, RLBP, CSLBP, FLBP, HOG, GLCM, Gabor and LGA using the LSVM classifier

Method	1/1 ROI		1/2ROI	
	Sensitivity	Specificity	Sensitivity	Specificity
LDN	**80.00**±0.1886	**96.00**±0.0699	64.00±0.1430	**99.00**±0.0316
LBP	77.00±0.1703	97.00±0.0483	57.00±0.1160	99.00±0.0211
RLBP	78.00±0.1135	97.00±0.0483	69.00±0.1729	99.50±0.0158
CSLBP	79.00±0.1287	93.00±0.0823	62.00±0.1398	98.50±0.0242
FLBP	21.00±0.1101	98.00±0.0632	41.00±0.2234	80.50±0.2034
HOG	72.00±0.0919	95.00±0.0527	63.00±0.1418	97.00±0.0422
GLCM	58.00±0.2394	63.00±0.1636	46.00±0.1578	79.00±0.2092
Gabor	63.00±0.1829	85.00±0.1354	59.00±0.1370	90.50±0.0497
LGA	55.00±0.1269	65.00±0.2550	42.00±0.1751	95.50±0.0369

Table 3.5: Sensitivity and specificity of breast mass detection with LDN, LBP, RLBP, CSLBP, FLBP, HOG, GLCM, Gabor and LGA using the NLSVM classifier (RBF kernel)

Method	1/1 ROI		1/2ROI	
	Sensitivity	Specificity	Sensitivity	Specificity
LDN	**81.00**±0.0738	**91.00**±0.1370	75.00±0.1581	**97.00**±0.0258
LBP	**85.00**±0.1179	**94.00**±0.0843	78.00±0.1317	**98.50**±0.0242
RLBP	**83.00**±0.0949	**89.00**±0.1197	**83.00**±0.1160	**91.50**±0.0747
CSLBP	78.00±0.1135	92.00±0.0789	64.00±0.1578	97.00±0.0422
FLBP	53.00±0.0823	75.00±0.1650	33.00±0.2003	88.50±0.2274
HOG	**81.00**±0.0876	**85.00**±0.1509	75.00±0.1434	**89.00**±0.0843
GLCM	67.00±0.1947	52.00±0.1549	58.00±0.2573	71.00±0.2025
Gabor	52.00±0.2658	71.00±0.2885	54.00±0.3438	60.50±0.3086
LGA	67.00±0.1337	49.00±0.1792	42.00±0.1398	95.00±0.0527

problems in describing breast tissues in dense breasts, which lowered their results. For example, one of the problems of the LBP, RLBP and FLBP descriptors is that

they may assign the same binary code for a pixel in the tumorous region and another pixel in normal dense tissue because they usually calculate the same difference between intensities. In addition, GLCM features depend on the co-occurrence matrix, which counts the number of the pixels that have the same intensity at a particular offset (distance and angle). Unfortunately, the same co-occurrence matrix can be provided for a tumorous ROI and a normal ROI in a dense breast region. Moreover, the problem of the HOG descriptor is the selection of the cell size and number of cells per block because if an unsuitable block size is used, the same HOG descriptor will be produced for a dense normal block and a tumorous block, leading to a high number of false detections.

3.4.2 Analysis of the feature combinations

On the basis of each texture analysis method we selected the ones with best specificity and sensitivity. From our experiments, we found that LDN, LBP, RLBP and HOG produced the best sensitivity and specificity with the NLSVM and RF classifiers. We tried to combine them to improve their individual accuracy. On the one hand, we concatenated the feature vectors of $LDN+LBP$, $LDN+RLBP$, $LDN+HOG$, $LBP+HOG$, $RLBP+HOG$, $LDN+LBP+HOG$ (named FC1), $LDN+RLBP+HOG$ (named FC2), and trained a new NLSVM or RF model in each case. We used *principal component analysis* (PCA) to reduce the dimensionality of the concatenated feature vectors (Jolliffe, 2005). We also used the classifier *majority voting* technique to combine the outputs of LDN, LBP, RLBP and HOG, using two combinations: $LDN+LBP+HOG$ (named MV1), and $LDN+RLBP+HOG$ (named MV2). The majority voting technique is applied within NLSVM models or RF models, trained with the aforementioned features, separately. Table 3.6 presents the results of the combination schemes in the case of the NLSVM classifier. Sensitivity is best with $LDN+LBP$, with both the balanced and imbalanced datasets. Specificity is 97% with FC1 with a balanced dataset, and 99% with $LDN+LBP$ with an imbalanced dataset. Moreover, Table 3.7 shows that $LDN+LBP$ gives good sensitivity and specificity with both datasets. $LDN+RLBP$ gives the best sensitivity with a

balanced dataset.

Table 3.6: Mean sensitivity and mean specificity of each feature combination strategy using the NLSVM classifier

Method	1/1 ROI		1/2ROI	
	Sensitivity	Specificity	Sensitivity	Specificity
$LDN + LBP$	**92.00**±0.0632	94.00±0.0516	**90.00**±0.0667	**99.00**±0.0211
$LDN + RLBP$	82.00±0.1476	93.00±0.0675	79.00±0.0994	98.50±0.0337
$LDN + HOG$	85.00±0.0850	94.00±0.0843	76.00±0.1350	97.50±0.0354
$LBP + HOG$	88.00±0.1135	93.00±0.0483	89.00±0.0876	97.50±0.0354
$RLBP + HOG$	84.00±0.0843	96.00±0.0516	79.00±0.1524	96.50±0.0530
$FC1$	86.00±0.0699	**97.00**±0.0483	78.00±0.1317	98.50±0.0337
$FC2$	82.00±0.1874	96.00±0.0516	76.00±0.1350	98.50±0.0242
$MV1$	84.00±0.0966	93.00±0.0675	82.00±0.1229	98.00±0.0258
$MV2$	83.00±0.1160	91.00±0.0738	87.00±0.0949	95.50±0.0438

Table 3.7: Mean sensitivity and mean specificity of each feature combination strategy using the RF classifier

Method	1/1 ROI		1/2ROI	
	Sensitivity	Specificity	Sensitivity	Specificity
$LDN + LBP$	88.00±0.1398	**96.50**±0.0516	**87.00**±0.0966	**98.50**±0.0425
$LDN + RLBP$	**90.00**±0.0816	96.00±0.0699	83.00±0.1252	97.50±0.0337
$LDN + HOG$	86.00±0.0516	91.00±0.1197	80.00±0.0816	97.50±0.0486
$LBP + HOG$	86.00±0.0966	91.00±0.0568	86.00±0.0966	97.50±0.0264
$RLBP + HOG$	87.00±0.1059	93.00±0.0823	84.00±0.0823	97.50±0.0354
$FC1$	85.00±0.1354	95.00±0.0850	83.00±0.0949	98.00±0.0258
$FC2$	88.00±0.1229	93.00±0.0675	81.00±0.0994	98.00±0.0258
$MV1$	86.00±0.1174	93.00±0.0823	84.00±0.1506	97.50±0.0425
$MV2$	83.00±0.0949	93.00±0.0823	82.00±0.1229	97.50±0.0354

Indeed, other combination methods ($FC2$, $MV1$ and $MV2$) gave lower sensitivities and specificities than $LDN + LBP$ and $LDN + RLBP$. The later combinations integrate the LDN descriptor (which depends on the minimum and maximum edge responses in the neighborhood of each pixel) with the LBP descriptor (which provides a description of all the elements in the neighborhood of each pixel). LBP focuses on encoding the variation of the grey levels in a local neighborhood, while LDN focuses on encoding the variation of the edge responses of each pixel in eight directions. In this way, we integrate the description abilities of LDN and LBP. Moreover, the use of the histograms of LDN and LBP leads to an overall description for each ROI.

Dividing a given ROI into sub-regions also adds some spatial information for the descriptors used. In this study, we did not discuss the effect of the rotation on the accuracy of the proposed methods because the mammograms are acquired from fixed positions (i.e., CC and MLO mammographic views). $LDN + LBP$ gives the smallest percentage of false positives; as a result, if it were used in a CAD system the number of unnecessary biopsies would be decreased.

3.5 Conclusion

In this chapter we have compared several texture analysis methods for breast mass detection, using images from a public breast cancer database. In particular, we have proposed LDN as a new feature extraction method in this field. LDN improved the results of such well-known texture analysis methods as LBP, HoG, GLCM or Gabor filters. In order to improve mass detection rates, we proposed using two combination schemes. First, we concatenated the features of the best texture analysis methods. Of all the concatenations, LDN+LBP gave the best overall results with the NLSVM and RF classifiers. Second, we used the classifier voting technique to combine the predictions given by LDN, LBP, RLBP and HOG. The results were good. The $LDN + LBP$ combination detected the false positives of breast mass detection; as a result, the number of unnecessary biopsies would be reduced (less diagnosis cost). In the next chapter we propose a CAD system to classify breast tissue into normal or mass and estimate the breast density. We propose the ULDP descriptor for feature extraction.

CHAPTER 4

Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern

4.1 Introduction

Mammograms are regarded as the best screening tool for detecting breast cancer early. The most common risk factors of breast cancer are *age*, *family profile*, *genetics* and *breast density*. Breast density is the amount of dense tissue in the breast, and it is the clearest risk factor of breast cancer (Lokate et al., 2010). The higher the breast density, the greater the probability of breast cancer. In addition, there is a relation between the age of women and their breast densities, as younger women

usually have denser breasts than older women. The breast masses are brighter than normal tissues in mammograms, and they are defined by their *shapes* and *margins*. In practice, mass detection is a big challenge in dense mammograms because normal tissues also appear as bright areas, and cover the places that contain masses. Some of the breast density standards classify breast tissues into *fatty*, *glandular* or *dense* (Suckling et al., 1994). Fig. 4.1 shows examples of these tissues in mammograms. It can be seen that it is easy to distinguish between mass and normal regions in the case of *fatty* tissues (the red circle refers to the location of the mass region), whereas mass detection is very difficult in the case of the *dense* mammogram. The *fatty-glandular* breasts are between these two cases. Furthermore, the breast imaging reporting and data system standard (BI-RADS) (Orel et al., 1999), presented by the American College of Cancer, provides the following breast density classification:

- BI-RADS I: almost entirely fatty breast (0-25%).
- BI-RADS II: some fibroglandular tissue (26%-50%).
- BI-RADS III: heterogeneously dense breast (51%-75%).
- BI-RADS IV: extremely dense breast (76%-100%).

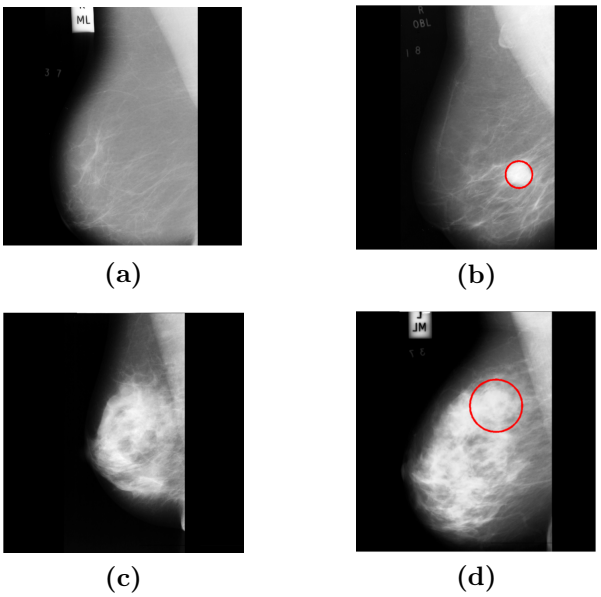


Figure 4.1: Examples of mammograms from the mini-MIAS breast cancer database (Suckling et al., 1994). A *fatty* mammogram containing: (a) normal tissue and (b) mass tissue. A *dense* mammogram containing: (c) normal tissue and (d) mass tissue

A breast cancer *computer-aided diagnosis* (CAD) system is a program that uses digitized (film-screen) or digital mammograms to assist radiologists by enhancing the quality of mammograms and detecting the early signs of breast cancer. Although radiologists try hard to estimate breast density and detect masses by making a visual judgment of mammograms, they insist on requesting CAD systems to help them in this difficult task.

The breast cancer CAD systems exploit various *computer vision* and *image processing* techniques. In general, a CAD system consists of three main steps: *segmentation* of the ROIs from the images, *feature extraction* from the ROIs and final *classification*. Both *breast density* classification and *breast mass* detection play an important role in improving of the treatment of breast cancer. Unfortunately, CAD systems do not perform as well in the case of dense mammograms (see Fig.4.1). Noise and artifacts in mammograms may also degrade the performance of feature extraction methods. The common artifacts that can exist in mammograms are: detector-based, machine-based, patient-related, processing and storage artifacts. The literature shows no consensus on the optimal set of features that characterize breast tissue. A poor description of breast tissues leads to a high number of false positives (mammograms interpreted by a CAD system as abnormal cases when they are actually normal).

The work reported in this chapter proposes a CAD system for analyzing breast tissues. This CAD system performs two tasks: breast tissue classification within a region of interest (mass or normal) and breast density classification. The main contributions of this chapter are the following:

- We propose the *uniform local directional pattern* (ULDP) as a texture descriptor for breast tissues in mammograms. ULDP codes a local neighborhood in the breast region based on its edge responses. As will be shown in this chapter, the proposed descriptor can discriminate between different masses in mammograms regardless of their size, shape or margin, and significantly improve the analysis of breast cancer.
- ULDP is used to classify breast tissues into mass or normal, and estimate breast

density. Two publicly available mammographic databases are used: mini-MIAS (screen-film database) and INbreast (full field digital database).

- We studied the effect of breast density on the performance of ULDP when classifying breast tissues into mass or normal.

4.2 Related work

This section summarizes some of the other studies that have been made on mass/normal breast tissue classification and breast density classification. We discuss the feature extraction methods used in each study and highlight the advantages of the descriptor proposed.

4.2.1 Breast tissue classification

Fig. 4.2 shows the common shapes and margins of breast masses. The shape of a particular breast mass can be *round*, *oval*, *lobular* or *irregular*. The circumscribed oval and round masses are usually *benign*, whereas *malignant* masses usually have irregular shapes. The margins of breast masses can be *circumscribed*, *microlobulated*, *obscured*, *indistinct* or *spiculated* (García-Manso et al., 2013).

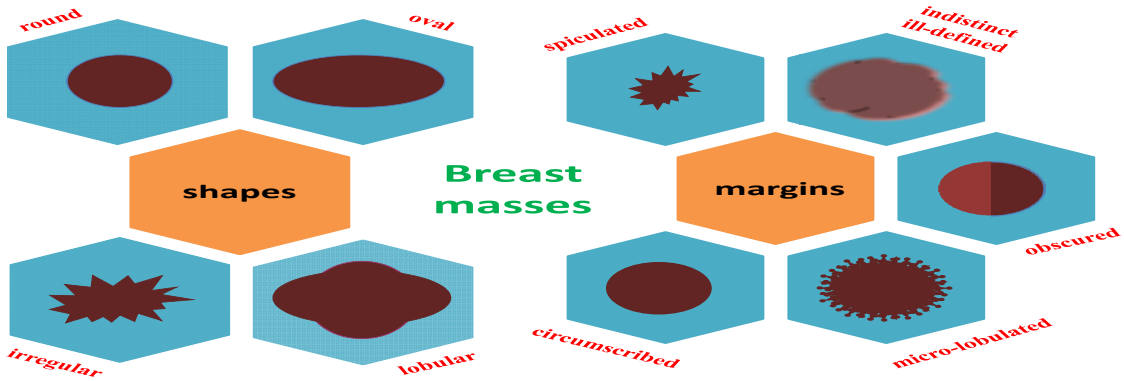


Figure 4.2: Common shapes and margins of breast masses

Several features extraction methods have been proposed for breast tissue classification. Here we present several studies on breast tissue classification, and discuss the descriptors they used. Table 4.1 summarizes some of this previous work.

Table 4.1: Summary of breast tissue classification methods

Method	Feature extraction method	Utilized classifiers
(Oliver et al., 2007)	LBP	SVM
(Liu and Zeng, 2015)	GLCM and completed local binary pattern	SVM
(Beura et al., 2015)	GLCM	Back propagation neural network
(Chu et al., 2015)	Shape features and texture features	SVM
(Jen and Yu, 2015)	Mean and standard deviation	Principal component analysis
(Wang et al., 2014b)	GLCM and morphological features	Extreme learning machine
(de Oliveira et al., 2015)	Taxonomic diversity index and taxonomic distinctness	SVM
(Dheeba et al., 2014)	Laws texture energy measures	Particle swarm optimized wavelet neural network
(Zheng, 2010)	Gabor filters	Threshold-based approach
(Pomponiu et al., 2014)	HOG	SVM

Oliver et al. (2007) used the LBP to reduce the number of false positives in breast mass detection. LBP can assign the same pattern to a pixel in a *tumorous region* and to another pixel in a *normal dense* tissue, which leads to a noticeable number of false detections, as illustrated in the section below. This happens when the values of all the neighbors are *higher/smaller* than the value of the central pixel. This problem of LBP is called the *saturation problem*.

The approaches that use the GLCM features still produce a noticeable number of false positives. The main reason for this is that the GLCM features basically depend on the co-occurrence matrix, which counts the number of pixels that have the same value at a particular offset (distance and angle). Thus, identical matrices are produced for the tumorous and normal ROIs in dense mammograms, thus increasing the number of false positives.

The shape, size and morphological features depend on the accuracy of the segmentation step, so segmentation errors degrade the performance of those descriptors. Moreover, the texture features used depend on the intensity values of the ROIs. Thus, the same features are computed for a dense normal pixel and a tumorous pixel, which causes a dilemma in the classification step. Although the taxonomic indices produce a good mass/normal classification result, they fail to classify mass and normal tissues in dense ROIs (de Oliveira et al., 2015).

Both HOG and LDN use *directional information* to describe breast tissues. In the

case of HOG, if the block size is unsuitable, the same HOG descriptor is produced for a normal block and a tumorous block in dense mammograms, which leads to a high number of false detections. Although LDN gives good results, it only encodes the maximum and minimum of the edge responses for each pixel, leading to a loss of relevant information about the neighborhood.

The aforementioned descriptors may fail with dense mammograms because they can not properly discriminate between mass and normal ROIs in dense mammograms. Noise and artifacts in mammograms (e.g. dead pixels, dead or unread lines, non-uniformities and ghosting (Ayyala et al., 2008)) may also degrade the performance of those feature extraction methods.

Unlike the proposals made in other studies, the proposed ULDP descriptor describes each neighborhood in a given ROI on the basis of its edge responses, taking into account both magnitudes and spatial information. It does not depend on the intensity values of the images. ULDP generates binary codes, from which we select the uniform patterns. These patterns are used to construct a histogram. The non-uniform patterns are assigned to a single bin in the ULDP histogram because they represent unwanted noise. As we will show in section 4.3.1, ULDP generates different histograms for mass and normal breast tissues even in dense mammograms. Although many CAD systems such as (Chu et al., 2015) use a segmentation method to determine the suspicious regions in mammograms, in this study we analyze breast tissue in manually selected ROIs to avoid the errors of segmentation methods especially in dense mammograms.

4.2.2 Breast tissue density classification

The relationship between breast density and the risk of developing breast cancer was reported by Wolfe in (Wolfe, 1976), where breast density was categorized into different classes. Nowadays, BI-RADS breast density categories are widely used. Below, we show several breast density classification methods and discuss the feature analysis methods they use.

Although such studies as (Zheng et al., 2015; Angulo et al., 2015; Muhimmah and

Zwiggelaar, 2006) have focused on the use of gray level histograms it seems that these histograms might not be sufficient for classifying mammograms into BI-RADS categories (Oliver et al., 2008). The main reason is that similar histograms may be produced for different breast densities.

Several descriptors have been used to analyze breast density: for example, LBP (Zheng et al., 2015), GLCM (Petroudi et al., 2015) and Gabor filters (Gamdonkar et al., 2015). In (Oliver et al., 2015), texture features (such as LBP, GLCM), intensity and morphologic features are fed into a SVM to distinguish between fatty and dense pixels. These descriptors may be affected by noise, so they cannot properly describe breast densities. The descriptor proposed in this chapter is more robust against the noise and changes in breast density, as will be shown below.

Several breast density classification methods have used a *segmentation step* to segment the breast region into different clusters. In (Oliver et al., 2005), each mammogram is divided into two clusters based on a fuzzy C-means algorithm. Then nine GLCM features are extracted from each cluster to train two classifiers: a k-nearest neighbor (KNN) classifier and a decision tree. The classification accuracies were 67% (KNN) and 73% (ID3). In (Oliver et al., 2008), the method proposed in (Oliver et al., 2005) is used to classify the breast density of mammograms retrieved from the mini-MIAS and DDSM databases. To avoid the errors of segmentation methods, several studies have proposed to extract features from manually selected ROIs. Indeed, radiologists generally focus on the breast region in each mammogram to be analyzed. In (Sharma and Singh, 2014), GLCM features, histogram-based features, fractal features and Law's texture energy measures are extracted from each ROI and classified into fatty or dense. In (Sharma and Singh, 2015; Muštra et al., 2012), the GLCM features are extracted from each ROI, while (Subashini et al., 2010) extracted various statistical features from each ROI to train a SVM.

The performance of the descriptors used in the methods discussed above may be degraded by the noise and artifacts in mammograms (e.g. dead pixels, dead or unread lines, non-uniformities and ghosting). The GLCM and LBP may produce similar descriptors for different tissues. The studies discussed above reveal that the

use of ROIs is one of the most appropriate techniques in breast tissue analysis. In this study, we extract ULDP features from manually selected ROIs and then we use a multi-class SVM to predict the density of each ROI.

4.3 Methods

The main steps in the proposed breast cancer CAD system are *ROI selection*, *feature extraction* and *classification*. A physician selects a ROI from the breast region to be analyzed by the CAD system. Given the selected ROI, ULDP features are extracted and sent to a classifier to perform two classifications: *mass/normal* tissue classification and *breast density* classification. SVMs have been widely used in breast tissue classification and breast density classification methods, and results have been good in both tasks. Therefore, we use a linear SVM and a non-linear SVM in the classification step.

4.3.1 Uniform local directional pattern

In this section we describe the most related descriptors (e.g. (Oliver et al., 2007; Ramirez Rivera et al., 2013)) and we provide a detailed explanation of the proposed ULDP descriptor.

Related descriptors. An important drawback of LBP is that it depends on the intensity difference of the pixels, which is very sensitive to noise and illumination changes. Fig. 4.3 presents an example of the calculation of LBP codes of two pixels: pixel A, which lies in a *tumorous* region, and pixel B, which lies in a *normal* region. Although A and B belong to completely different regions, LBP assigns *00000000* to both pixels, which creates a dilemma in the classification step.

Chen et al. proposed a method for correcting the non-uniform patterns in the LBP binary codes to reduce the effect of noise. Chen's method partitioned each 8-bit LBP binary code into sequences of three consecutive bits. If *010* or *101* are found in the binary code, they are replaced by 000 or 111 respectively. Chen's method converts a natural non-uniform pattern into a uniform pattern, which distorts the overall

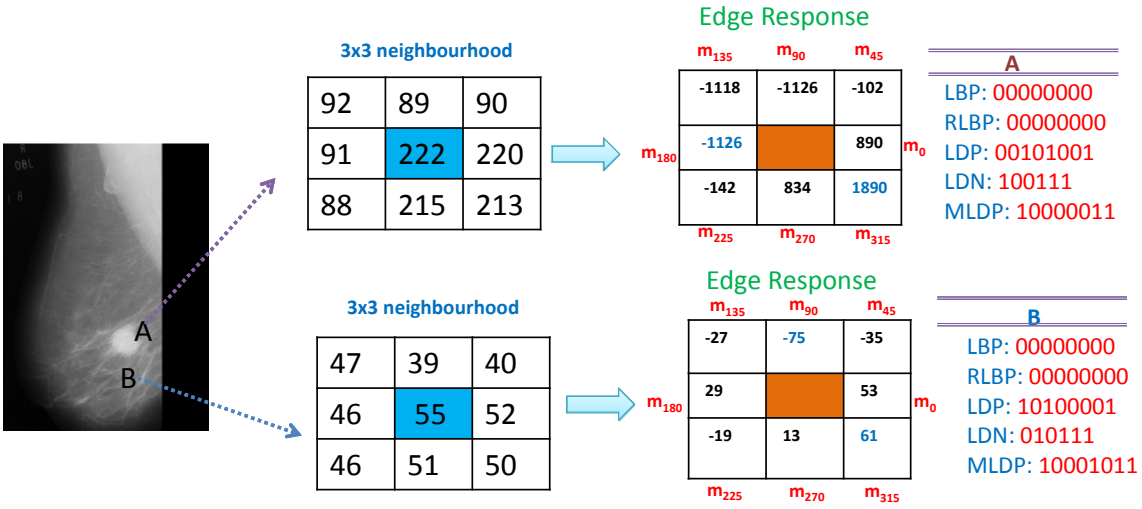


Figure 4.3: The binary codes generated with LBP, RLBP, LDP, LDN and MLDP for pixel A, which belongs to a *tumorous* region, and pixel B, which belongs to a *normal* region

description, because all wrong corrections affect the quality of the final histogram.

Local directional pattern (LDP) is a robust texture descriptor that encodes the directional information (i.e., the edge responses) in a local neighborhood. Therefore, LDP reduces the LBP dependence on the intensity difference.

LDP computes the edge responses in eight directions by using the compass Kirsch masks (Gonzalez and Woods, 2002). Jabid et al. (2010) set the positions of the top *three* responses to 1 and the positions of other directions to 0, generating an 8-bit binary code for each pixel. Fig. 4.3 shows an example of the calculation of the LDP codes for pixels A and B. LDP assigns *00101001* to pixel A and *10100001* to pixel B. Although LDP generates different codes, the reliance on only the top three responses leads to a loss of information about the local neighborhoods.

An example of LDN calculation is presented in Fig. 4.3. LDN can distinguish between pixels A and B by assigning different binary codes *100111* and *010111*, respectively. However, this descriptor also leads to a loss of information about the neighbors in a particular neighborhood, because only the minimum and maximum responses in the neighborhood are considered. In (Mohamed et al., 2014), a *modified local directional pattern (MLDP)* is proposed for an illumination-robust optical flow estimation. MLDP encodes the edge responses computed by using compass masks based on a Gaussian filter. In MLDP, the positive responses are set to 1 and the

bits corresponding to negative responses are set to 0. As shown in Fig. 4.3, MLDP produces 10000011 and 10001011 for pixels A and B, respectively. Although MLDP can discriminate between the two pixels, it is still sensitive to noise because there are many *bitwise jumps* in their binary codes.

Proposed descriptor. We propose the *uniform local directional pattern* (ULDP) as a texture descriptor for breast tissues in mammograms. ULDP describes each neighborhood in the breast region on the basis of its edge responses, taking into account not only their magnitudes but also spatial information. ULDP computes the edge responses using the eight Kirsch compass masks (Gonzalez and Woods, 2002), and then all the positive responses are set to 1 (the intensity has changed from dark to light), whereas the negative responses are set to 0 (the intensity has changed from light to dark). Fig. 4.4 shows an example of ULDP code calculation for a pixel in a ROI.

Given an $N \times N$ gray level ROI, an 8-bit code is calculated for each pixel. After all the patterns in the ROI have been calculated, only the uniform patterns are considered. A *uniform pattern* contains at most two transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. The *non-uniform* patterns usually have some undesirable characteristics, such as partial correlation and unwanted noise (Ojala et al., 2002). For instance, 00111100 is a uniform pattern, whereas 00101010 is a non-uniform pattern. In ULDP, a separate label of the final histogram is set for each uniform pattern. The uniform mapping produces 58 output labels, whereas the other 198 non-uniform patterns are assigned to a single label. Thus, there are 59 labels in the resulting histogram.

Given an $N \times N$ gray level ROI, the ULDP codes are converted to decimal values and then the histogram of ULDP is computed as follows:

$$H(ROI) = \sum_{x,y} M\left(I(x,y) = i\right), \quad i = 0, 1, \ldots, n-1 \tag{4.1}$$

where I is a matrix containing the decimal values and n is the number of different labels (n=59). If $I(x,y)$ is equal to i then $M\left(I(x,y) = i\right) = 1$, otherwise

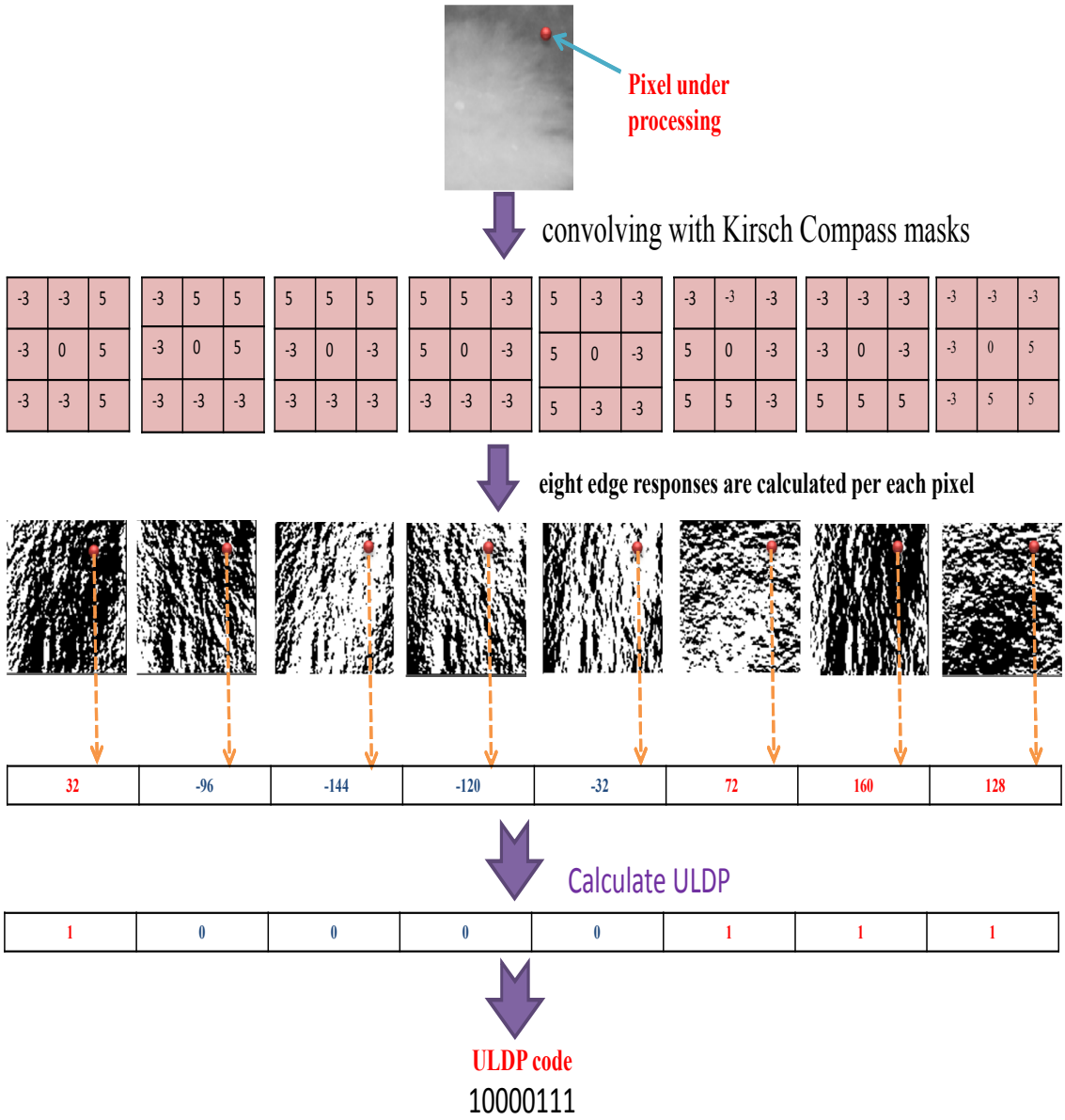


Figure 4.4: Example of ULDP code generation

$M\left(I(x,y)=i\right)=0$.

Fig. 4.5 shows how ULDP can distinguish between mass and normal breast tissues. Both normal and tumorous ROIs in a dense breast have similar visual characteristics, as can be observed in Fig. 4.5 (a) and Fig. 4.5 (b). LBP generated a large number of repetitions of the uniform pattern *11111111*, which is labeled in the bin number 58 of the LBP histogram, as shown in Fig. 4.5 (c) and Fig. 4.5 (d). This is the saturation problem of the LBP descriptor that was explained above. RLBP assigned the code 00000000 to many pixels in the normal and mass regions, a result of wrong corrections being made to the descriptor (many sub-patterns were wrongly converted from 010 to 000). In turn, ULDP produces distinctive histograms for the ROIs, and the resulting histograms are well distributed, as shown in Fig. 4.5 (g) and Fig. 4.5 (h). To increase the discriminative capability of ULDP, spatial information is considered by dividing each ROI into N_s subregions. The final descriptor H_{ULDP} of a given ROI can be constructed by concatenating the ULDP histograms as follows:

$$H_{ULDP}(ROI) = \biguplus_{j} H_j, \quad j = 1, 2, \ldots, N_s \tag{4.2}$$

where $\biguplus$ is the concatenation operator, H_j is the histogram of the subregion j, and N_s is the number of subregions.

In the following section we calculate the number of subregions and the optimal ROI size which yield the best classification result. ULDP can properly discriminate between different tissues in mammograms regardless of their *sizes*, *shapes* or *margins*, as shown in the experimental section below.

4.3.2 Classification

In this study we used LSVM and NLSVM with RBF to discriminate between mass and normal tissues. In the case of mass/normal breast tissue classification, the optimal pair (C,γ) of ULDP with the mini-MIAS database was (0.5, 0.015625) while with the INbreast database it was (0.25, 0.0625). In addition, a *multi-class classifier* based on a SVM was implemented to separate breast tissue densities. The

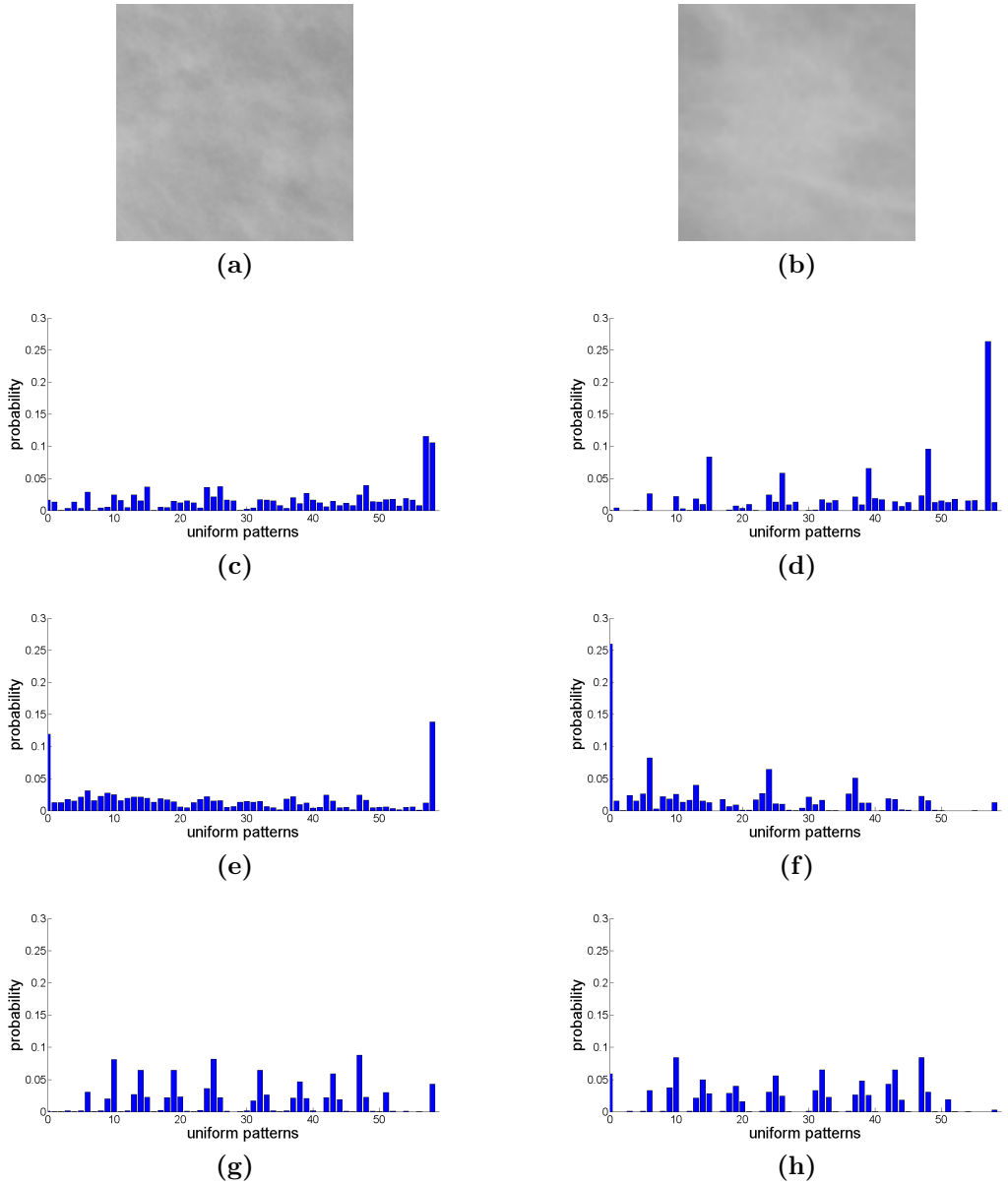


Figure 4.5: Comparison between the histograms of ULDP, LBP and RLBP. The first row shows two ROIs extracted from the mini-MIAS database: *normal ROI* (extracted from mdb174) and *mass ROI* (extracted from mdb204). The remaining figures show the resulting histograms of (a, b) with LBP (c, d), RLBP (e, f) and ULDP (g,h), respectively

common way to extend a binary SVM classifier to the multi-class scenario is to decompose an *M-class* problem into a series of binary problems. A set of binary classifiers are constructed to obtain the M-class classifier. Two different techniques can be applied: one-against-all and one-against-one. In this study, we used the one-against-all technique (Liu and Zheng, 2005; Hsu and Lin, 2002), where each classifier is trained to separate one class from the rest.

4.4 Experimental results

In this section the performance of ULDP in *mass/normal* breast tissue classification and *breast tissue density* classification is evaluated. Firstly, for *mass/normal breast tissue classification*, the ROIs were extracted from each database according to its ground truth (gold standard) (Pomponiu et al., 2014; Oliver et al., 2007). Then, the resulting histogram of each ROI was sent to a recognition system based on supervised learning: LSVM and NLSVM. A quantitative comparison between the classification with ULDP and some of the state-of-the-art descriptors, such as LBP (Oliver et al., 2007), HOG (Pomponiu et al., 2014), GLCM (Liu and Zeng, 2015; Beura et al., 2015) and Gabor filters (Zheng, 2010) was also made. Secondly, for the *breast tissue density classification*, the ROIs were extracted from the breast regions behind the nipple, as described in (Sharma and Singh, 2014; Muštra et al., 2012; Subashini et al., 2010). A multi-class classifier was used to classify each ROI into different breast densities. To show the robustness of the SVM in these two classification tasks we also evaluated the performance of two more classifiers: LDA and MLP. In the case of MLP, we used the configuration presented in (García-Manso et al., 2013). We used two breast cancer databases in our experiments: mini-MIAS (Suckling et al., 1994) and INbreast (Moreira et al., 2012).

The performance of ULDP in mass/normal breast tissue classification was measured in terms of the AUC of the ROC, whereas its performance in breast tissue density classification was measured in terms of the accuracy and the confusion matrix. The k-fold cross validation technique was used to generate the training and testing

datasets. In addition, the confusion matrix shows the TP, TN, FP and FN of the classification process on an $m \times m$ matrix, where m is the number of the classes. The diagonal of the confusion matrix shows the correctly classified instances. Furthermore, we used the *Kappa* statistic to measure the observed agreement among the raters of breast tissue density classification. This statistic compares the accuracy of the system to the accuracy of a random system. It is used to assess the inter-rater reliability when observing categorical variables (κ= 1 if there is full agreement; κ= 0 if there is no agreement).

4.4.1 Mass/normal breast tissue classification

In this section two experiments were made to evaluate the performance of ULDP when classifying mass and normal breast tissues. The first experiment used a set of ROIs extracted from the mini-MIAS database, while the second used a set of ROIs extracted from the INbreast database. Given a set of mass and normal ROIs, in both experiments the ULDP features were extracted from each ROI and sent to a LSVM or a NLSVM to be classified into mass or normal tissue. In order to create the ROIs, we followed the procedure given in (Pomponiu et al., 2014; García-Manso et al., 2013; Oliver et al., 2007). The GT of each database was used to determine the smallest square region that contains the mass. In turn, the ROIs of the normal tissues were randomly selected from the normal mammograms, and they were created with random sizes ranging from the smallest to the largest size found in the database.
Using the methodology explained above, 109 mass ROIs were extracted from the mass mammograms and 203 normal ROIs were extracted from the normal mammograms of the mini-MIAS database. In the case of the INbreast database, 107 mass ROIs were extracted from the mass mammograms and 300 from the normal mammograms. The extracted ROIs were of different sizes, so they were resized into a fixed template, keeping the same aspect ratio. It has been demonstrated that resizing the ROIs into a common template preserves the mass malignancy information (Campanini et al., 2004; Hong and Brady, 2003). To determine the optimal size of the ROIs, different sizes were considered (32×32, 64×64, 75×75 and 150×150 pixels) and the AUC

values for each size were computed. Fig. 4.6 shows the effect of the ROI size on the classification of breast tissue. The AUC values are plotted against the change in ROI size (the AUC values were calculated using LSVM). In the case of the mini-MIAS database, the best ROI size is 75×75 pixels, while the 64×64 and 75×75 sizes gave comparable AUC values in the case of the INbreast database. Thus, we recommend using a template with a size of 75×75 pixels.

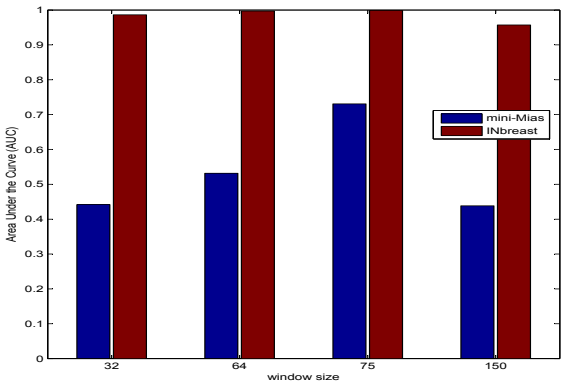


Figure 4.6: The effect of the ROI size on the performance of mass/normal breast tissue classification in the mini-MIAS (blue) and INbreast (red) databases

To add spatial information to ULDP, each ROI was divided into N_s subregions. The ULDP histogram was calculated for each subregion, and then the resulting histograms were concatenated to build the final descriptor (see section 4.3.1). The descriptor was then normalized to unit length using the L2-norm. Fig. 4.7 shows the relationship between the number of subregions and the resulting AUC values (calculated using LSVM). A total of 16 subregions yielded the highest AUC value with the mini-MIAS database and 4 subregions yielded the best AUC value with the INbreast database. The reason for the considerable variation in the number of subregions is that the INbreast database contains high contrast digital mammograms, so the masses and other internal structures in a breast region are more visible in INbreast than in mini-MIAS.

Each mammogram may contain many mass regions. Thus, if a normal/mass ROI of a certain mammogram is used in the testing stage, the other ROIs extracted from the same mammogram must be excluded in the training stage. This step ensures a

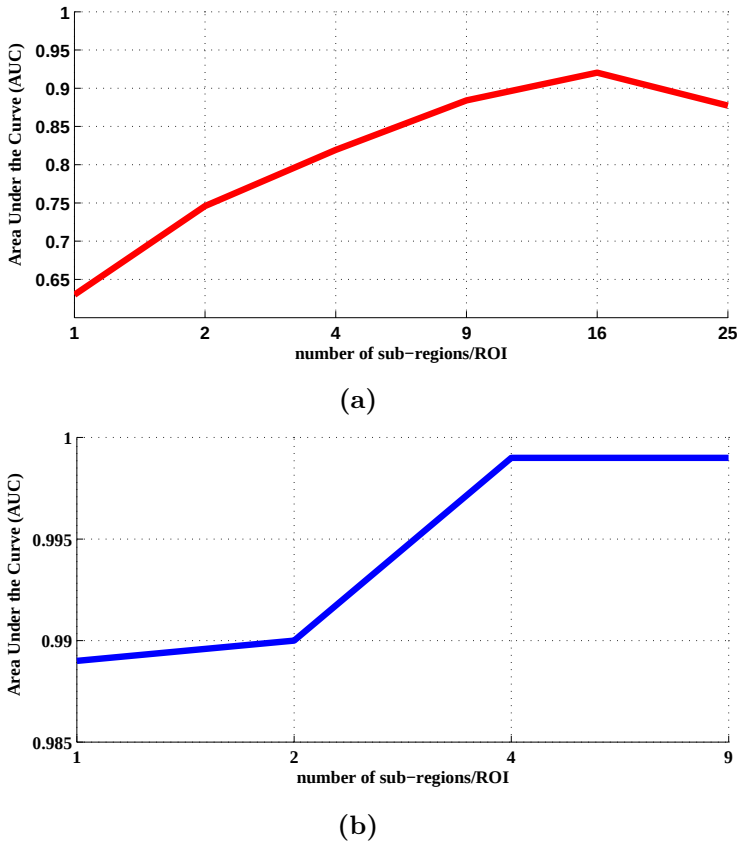


Figure 4.7: The effect of the number of subregions on mass/normal breast tissue classification with two databases: (a) mini-MIAS and (b) INbreast

full separation between the training and testing instances even when using a cross validation process.

Table 4.2 presents the classification results with the ULDP descriptor applied to the mini-MIAS database, as well as the results of mass/normal breast tissue classification with LBP (Oliver et al., 2007), RLBP (Chen et al., 2013), HOG (Pomponiu et al., 2014), LDP (Ramirez Rivera et al., 2013), MLDP (Mohamed et al., 2014), Gabor filters (Zheng, 2010) and GLCM (Liu and Zeng, 2015; Beura et al., 2015). All descriptors were normalized to unit length. We implemented LBP, RLBP, LDP and MLDP using the same procedure followed with ULDP (i.e., each ROI was divided into subregions and their histograms were concatenated). The values of the parameters in our implementation of HOG were a 3×3 cell size, 8×8 cells for the block size and a 9-bit histogram.

The results with ULDP outperformed all other descriptors, yielding an AUC value

of 0.9325 with the LSVM and 0.9292 with the NLSVM classifier. MLDP, Gabor filters and GLCM gave AUC values lower than 0.70 with LSVM and NLSVM, which show they are weak at describing the internal structures of the breast tissues in the case of the mini-MIAS database. LBP produced acceptable results with the NLSVM classifier, but the LBP features could not be linearly separated. RLBP gave good AUC values with both the LSVM and NLSVM classifiers.

Table 4.3 shows the results of the classification with the INbreast database. The AUC values obtained with ULDP, LBP, RLBP, HOG, LDP and GLCM were over 0.978 with LSVM and NLSVM. The main reason for these excellent results is that INbreast (which is a digital database) has very good contrast (i.e. masses are more visible in digital mammograms). LDP produced a slightly greater AUC value than ULDP because we added spatial information for LDP as we did in ULDP. The MLDP descriptor produced an AUC of 0.5254 with LSVM and an AUC of 0.7896 with NLSVM, indicating that these features cannot distinctively characterize breast tissues in digital mammograms.

Table 4.2: Comparison between the AUC values of mass/normal breast tissue classification using ULDP, LBP, RLBP, HOG, LDP, MLDP, Gabor and GLCM features using LSVM, NLSVM, LDA and MLP classifiers with the mini-MIAS database

Methods	LSVM	NLSVM	LDA	MLP
ULDP	**0.9325 ± 0.0630**	**0.9292 ± 0.0567**	**0.9133±0.0524**	**0.9224±0.0626**
LBP	0.6978 ± 0.0295	0.8947 ± 0.0159	0.7728±0.1349	0.6271±0.0722
RLBP	0.9103 ± 0.0697	0.9228 ± 0.0725	0.9277±0.0680	0.9295±0.0475
HOG	0.7664 ± 0.0311	0.8874 ± 0.0920	0.7495±0.1668	0.9065±0.0674
LDP	0.8195 ± 0.1310	0.705 ± 0.1935	0.7070±0.1134	0.7879±0.0828
MLDP	0.5404 ± 0.2150	0.5338 ± 0.1767	0.6258±0.0784	0.5397±0.0955
Gabor	0.6901 ± 0.1052	0.6412 ± 0.1546	0.8582±0.0763	0.7459±0.0870
GLCM	0.6803 ± 0.2154	0.6217 ± 0.2968	0.8490±0.0846	0.7884±0.1082

The statistical significance of the AUC values gives further evidence supporting ULDP. The statistical significance simply means that a given result is unlikely to have occurred by chance. The *t-student* test (David and Gunnink, 1997) was used to calculate the statistical significance of the AUC values. An AUC value of 0.5 means that the output is produced by chance. The number of experiments equals 8 and the significance level α of a single experiment equals 0.05; so, according to the Bonferroni correction (Curtin and Schulz, 1998), the actual significance level is

Table 4.3: Comparison between the AUC values of mass/normal breast tissue classification using ULDP, LBP, RLBP, HOG, LDP, MLDP, Gabor and GLCM features using LSVM, NLSVM, LDA and MLP classifiers with the INbreast database

Methods	LSVM	NLSVM	LDA	MLP
ULDP	**0.9930 ± 0.0102**	**0.9926 ± 0.0113**	**0.9413±0.0587**	**0.9987±0.0032**
LBP	0.9789 ± 0.0238	0.9918 ± 0.0104	0.9732±0.0577	0.9207±0.0595
RLBP	0.990 ± 0.0228	0.9902 ± 0.0022	0.9105±0.0631	0.9888±0.0148
HOG	0.9925 ± 0.0335	0.9934 ± 0.0293	0.9600±0.0459	0.9983±0.0050
LDP	0.9983 ± 0.0028	0.9978 ± 0.0045	0.8590±0.0831	0.9786±0.0189
MLDP	0.5254 ± 0.1292	0.7896 ± 0.1383	0.8402±0.0973	0.6080±0.1284
Gabor	0.9547 ± 0.0409	0.9468 ± 0.0504	0.2210±0.1170	0.4816±0.0854
GLCM	0.9875 ± 0.0035	0.9845 ± 0.0017	0.9898±0.0216	0.9979±0.0067

0.0063 (0.05/8). With the mini-MIAS database, the AUC values with ULDP, LBP, RLBP, HOG, LDP, Gabor filters and GLCM were statistically significant (p-value < 0.0001). In turn, the AUC values with MLDP were not statistically significant (p-value > 0.0063), indicating that the AUC values with the MLDP were produced by chance. With the INbreast database, the AUC values with ULDP, LBP, RLBP, HOG, LDP, MLDP and GLCM were statistically significant (p-value < 0.0001), but the AUC values with Gabor filters were not statistically significant (p-value > 0.0063).

Moreover, with the mini-MIAS database the AUC values with ULDP were statistically more significant (p-value < 0.0001) than the ones with LBP, MLDP and Gabor filters with LSVM; in addition, the AUC values with ULDP were statistically more significant than the ones with all the other descriptors with NLSVM. In the case of the INbreast database, considering both LSVM and NLSVM, the results with ULDP were statistically more significant than the results of all other descriptors. The overall results confirm that ULDP produces a good description of breast tissues.

In our experiments, we evaluated how other classifiers such as LDA and MLP classified mass/normal breast tissue using the proposed descriptor. In the case of the mini-MIAS database, LDA and MLP gave lower AUC values than with ULDP and LSVM. In contrast, RLBP with LDA and MLP gave better AUC values than with ULDP (Table 4.2). With the INbreast database, LDA and MLP with ULDP gave similar AUC values to those with LSVM and NLSVM (Table 4.3). LDA with LBP, HOG and GLCM gave higher AUC values than with ULDP.

We also calculated the statistical significance of the AUC values with LDA and MLP. With LDA, the AUC values of ULDP were statistically more significant than those of LDP and MLDP with the mini-MIAS database (p-value < 0.0063), while the AUC values of ULDP were statistically more significant than those of Gabor filters with the INbreast database. In the case of MLP, the AUC values of ULDP were statistically more significant than the ones of LBP, LDP, MLDP and Gabor filters with the mini-MIAS and the INbreast. The AUC values of ULDP were statistically more significant than the ones of GLCM with the mini-MIAS while they were not statistically significant with INbreast.

4.4.2 Breast tissue density classification

In this section we evaluated how well the ULDP descriptor classified breast tissue density. To classify the mammograms according to their breast densities, a 100×100 ROI of each mammogram was manually extracted as proposed in (Sharma and Singh, 2014; Muštra et al., 2012). The selected ROI contains breast tissues that lie behind the nipple, in the densest region of the breast (excluding the pectoral muscle). Each ROI was then divided into 25 subregions (in this study, this number was obtained empirically). Afterwards, the ULDP histograms were individually extracted from each subregion and concatenated to build the final descriptor. As explained in section 9.4.2, a multi-class supervised learning classifier based on an SVM was used to classify the ROIs into different densities.

Two experiments were performed with the mini-MIAS database: *fatty/dense* and *fatty/glandular/dense* breast tissue classifications. Table 4.4 presents the confusion matrix of the *fatty/dense* classification of 30 mammograms per class in which an accuracy of 96.67% with a Kappa coefficient $\kappa = 0.933$ was achieved. Table 4.5 presents the confusion matrix of the *fatty/dense* classification of 50 mammograms per class in which an accuracy of 98.0% ($\kappa = 0.99$) was obtained. The high accuracy of the results and the noticeable agreement between the raters attest the ability of ULDP to distinguish between fatty and dense tissues.

We also tested the classification into *fatty/glandular/dense* breast densities. The

Table 4.4: Confusion matrix of the *fatty/dense* breast tissue density classification in mini-MIAS

	fatty	dense
fatty	0.9667	0.0333
dense	0.0333	0.9667

Table 4.5: Confusion matrix of the *fatty/dense* breast tissue density classification in mini-MIAS

	fatty	dense
fatty	0.98	0.02
dense	0.02	0.98

glandular class is between fatty and dense tissues (i.e. it has characteristics in common with fatty and dense tissues). As shown in Table 4.6, ULDP achieved an accuracy of 87.78% with 30 mammograms per class ($\kappa = 0.8167$) and an accuracy of 88.0% ($\kappa = 0.820$) with 50 mammograms per class (Table 4.7). The correct classification rate in the *fatty/glandular/dense* classification is lower than the correct rate in the *fatty/dense* classification due to the addition of the glandular class.

Table 4.6: Confusion matrix of the *fatty/glandular/dense* breast tissue density classification in mini-MIAS

	fatty	glandular	dense
fatty	0.8333	0.1667	0
glandular	0.1	0.8	0.1
dense	0	0	1

Table 4.7: Confusion matrix of the *fatty/glandular/dense* breast tissue density classification in mini-MIAS

	fatty	glandular	dense
fatty	0.84	0.16	0
glandular	0.16	0.8	0.04
dense	0	0	1

The INbreast database contains only 28 BI-RADS IV mammograms. So, 30 mammograms from BI-RADS I, BI-RADS II and BI-RADS III were added, in an attempt to make a balanced classification. As shown in Table 4.8, accuracy was 92.37% and agreement high (κ=0.9091). This high accuracy certifies the ability of ULDP to discriminate between different breast densities. Table 4.9 shows the confusion matrix of an imbalanced classification process, in which 50 mammograms were used from BI-RADS I, BI-RADS II and BI-RADS III, as well as the 28 mammograms of BI-RADS IV. As a result of the imbalanced dataset, the accuracy decreased to 76.97%.

The performance of LDA and MLP with breast density classification was also evaluated. With mini-MIAS, MLP produced classification results similar to those produced with SVM. In the case of *fatty/dense* classification, MLP with the

Table 4.8: Confusion matrix of the breast tissue density classification in INbreast

	I	II	III	IV
I	0.9667	0.0333	0	0
II	0.0667	0.8667	0.0667	0
III	0	0	0.9333	0.0667
IV	0	0	0.0357	0.9643

Table 4.9: Confusion matrix of the breast tissue density classification in INbreast

	I	II	III	IV
I	0.82	0.1	0.04	0.04
II	0.14	0.7	0.1	0.06
III	0.06	0.04	0.86	0.04
IV	0	0.0714	0.1071	0.8214

mini-MIAS achieved an accuracy of 96.66% with 30 mammograms per class ($\kappa = 0.93$) and an accuracy of 97.67% ($\kappa = 0.9733$) with 50 mammograms per class. In the case of *fatty/glandular/dense* classification, MLP with the mini-MIAS achieved an accuracy of 78.24% with 30 mammograms per class ($\kappa = 0.656$) and an accuracy of 81.3% ($\kappa = 0.7184$) with 50 mammograms per class. The classification results of MLP were much lower than those of SVM in the case of INbreast. It achieved an accuracy of 26.67% ($\kappa = 0.0248$) with 30 mammograms per class and an accuracy of 39.26% ($\kappa = 0.1655$) with 50 mammograms per class. LDA was much less accurate than SVM in the case of mini-MIAS and INbreast (max. 50%).

4.4.3 Effect of breast density on the performance of ULDP

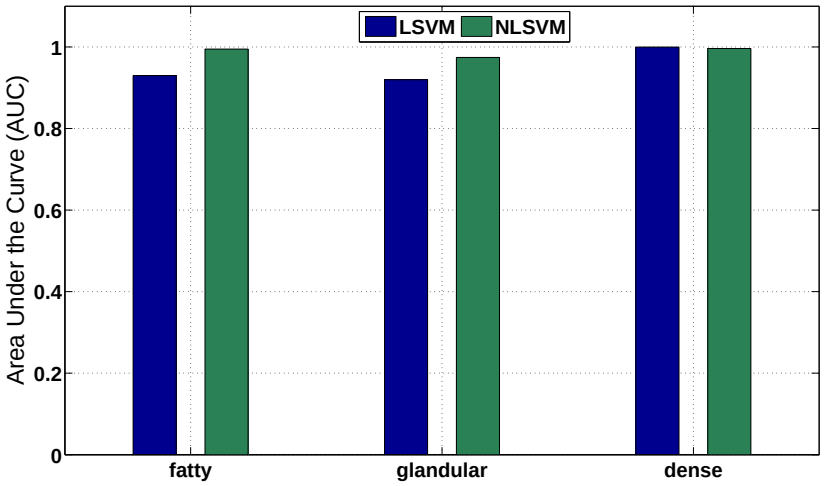
The effect of breast density on the performance of mass/normal breast tissue classification using ULDP is at the heart of this discussion. In practice, the higher the breast density, the more difficult it is to detect the breast masses. The effect of the breast density on mass/normal breast tissue classification was discussed in (García-Manso et al., 2013) using a set of ROIs extracted from the DDSM database (Heath et al., 1998). In this chapter we studied the effect of breast density on the ability of ULDP to perform *mass/normal* breast tissue classification by considering the mass and normal ROIs extracted from the same breast density class. For instance, in order to study the effect of the *fatty tissue*, the classification process was performed using the mass and normal ROIs extracted from *fatty mammograms*. In the case of mini-MIAS, we divided the mammograms into *three datasets* according to their breast densities (fatty, glandular and dense). Consequently, three datasets were generated: *fatty* (106 mammograms), *glandular* (104 mammograms) and *dense*

(112 mammograms). In each dataset, the mass and normal mammograms were identified. The *mass/normal* classification was then performed using these extracted *mass* and *normal* ROIs (the ROIs were generated using the methodology explained in section 4.4.1). In turn, the INbreast database was divided into *four datasets*: BI-RADS I (132 mammograms), BI-RADS II (148 mammograms), BI-RADS III (97 mammograms) and BI-RADS IV (28 mammograms). On the basis of a leave-one-out cross validation study, Fig. 4.8 presents the effect of breast density on the ability of ULDP to classify breast tissues into normal or mass. Both LSVM and NLSVM classifiers were used in the classification step. With the mini-MIAS database, *mass/normal* breast tissue classification using NLSVM performs more or less constantly with all densities. However, there is a small variation in the AUC values in the case of LSVM. In the case of the INbreast database, there is a small variation in the AUC values with LSVM and NLSVM. These experiments attest that ULDP performs well and indicates that it is able to characterize different tissues in the breast area, regardless of density.

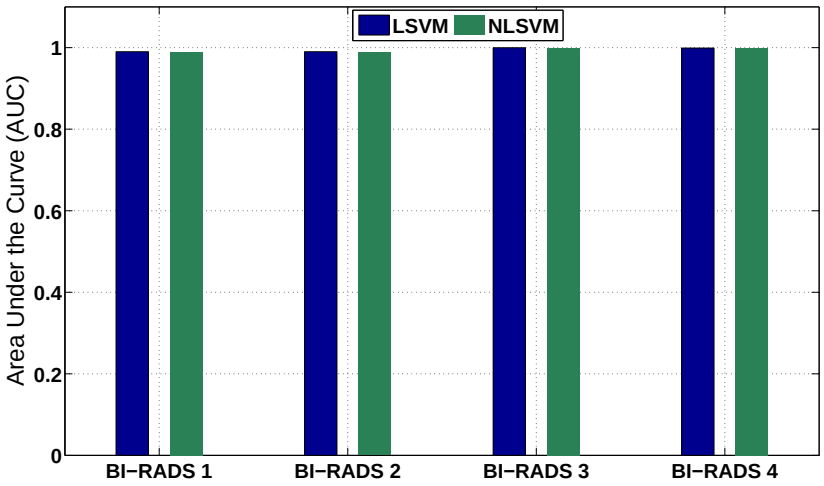
4.5 Discussion

The main advantage of ULDP is that it uses uniform patterns to build a robust descriptor. The non-uniform patterns are assigned to a single bin in the final histogram because they usually have undesirable characteristics, such as partial correlation and unwanted noise. Moreover, calculating the edge responses in eight different directions provides a good representation for the micro-patterns in a given ROI. Consequently, if a micro-pattern is missed in a particular direction, it can be captured in other directions. In this way, ULDP can describe the structures of different tissues in the breast region.

As observed from the experiments, the optimal number of subregions varies in the film-screen and FFDM databases, because the breast tissues are more visible in the digital mammograms, whereas a film-screen mammogram usually contains blurred regions and noise in the breast area (Pisano et al., 2005). The ROIs generated were of



(a)



(b)

Figure 4.8: The effect of breast density on the discrimination ability of ULDP with LSVM and the NLSVM and two databases: (a) mini-MIAS and (b) INbreast

different sizes, because the masses are naturally of different sizes. Thus, all the ROIs were resized into a template with a common size. With the mini-MIAS database, the best ROI size in mass/normal breast tissue classification was 75×75 pixels, whilst both 64×64 and 75×75 pixels gave comparable results in the case of INbreast (in this study, the template size was 75×75 pixels). ULDP outperformed some of the state-of-the-art descriptors in *mass/normal* breast tissue classification such as LBP (Oliver et al., 2007), HOG (Pomponiu et al., 2014), GLCM (Liu and Zeng, 2015; Beura et al., 2015) and Gabor filters (Zheng, 2010). In addition, ULDP gave the best AUC (0.9325) with the mini-MIAS database and its results were statistically significant, which means that ULDP provides a distinctive description for breast masses in film-screen mammograms. In the case of INbreast, results were good with ULDP, LBP, RLBP and LDP (about 0.99) as a result of the clear visibility of the breast tissues in digital mammograms.

In addition, the ULDP descriptor was evaluated for its ability to classify breast tissue density. With the mini-MIAS database, results were reliable in both *fatty/dense* and *fatty/glandular/dense* classification. The main reason for these results is that ULDP is independent of the *intensity values*, because it is a structural descriptor based on edge responses. Table 4.10 presents a quantitative comparison between the accuracy of breast tissue density classification using ULDP and the results of the methods proposed in (Oliver et al., 2008, 2005; Muštra et al., 2012; Sharma and Singh, 2014) which were applied to the *mini-MIAS* database. The working principles of these methods have been discussed in section 4.2.2.

Table 4.10: Comparison between *fatty/dense* and *fatty/glandular/dense* breast density classification using ULDP with some related studies which used the mini-MIAS database

Method	2 classes	3 classes
Proposed	**99.0%**	**85.5%**
Oliver (Oliver et al., 2008), (Oliver et al., 2005)	91%	73%
Mustra (Muštra et al., 2012)	91.6%	82.5%
Sharama (Sharma and Singh, 2014)	96.46%	–

The proposed method gave better accuracy than other methods in both *fatty/dense* and *fatty/glandular/dense* breast tissue density classification. In addition, the classification of breast tissue density with the proposed method based on ULDP achieved high agreement between raters ($\kappa = 0.82$), while (Oliver et al., 2008) achieved a kappa coefficient of 0.81.

With the INbreast database, breast density was classified into four classes. Only a few studies have used digital databases, so a quantitative comparison was made with only two related studies (Tortajada et al., 2012; Muštra et al., 2012). Tortajada et al. used a private FFDM database for breast density classification and they achieved an accuracy of 92.0% with $\kappa = 0.88$. Muštra et al. achieved an accuracy of 76.2% with a private FFDM database.

Table 4.11 compares the classification accuracy of the proposed method based on ULDP and other methods (Muštra et al., 2012; Tortajada et al., 2012) when they are applied to the INbreast database. Our approach achieved an accuracy of 92.37% with $\kappa = 0.9091$, higher than that of other methods.

Table 4.11: Comparison of the accuracy of breast density classification of ULDP with (Muštra et al., 2012) and (Tortajada et al., 2012) using INbreast

Method	Accuracy
Proposed	**92.37%**
Mustra (Muštra et al., 2012)	88.2%
Tortajada (Tortajada et al., 2012)	89.0%

In this study we used LSVM and NLSVM to classify breast tissues into mass or normal and to classify breast density. We also evaluated the performance of LDA and MLP with the proposed ULDP descriptor. For breast tissue classification, LSVM and the NLSVM gave the highest AUC values with the mini-MIAS database, while all the classifiers gave comparable AUC values with the INbreast database. In the case of breast density classification, the multi-class SVM was most accurate with the mini-MIAS and INbreast databases.

Our experiments reveal that ULDP can properly describe breast tissues even in dense mammograms. Although the proposed descriptor gave good results for breast tissue

classification and breast density classification, the main limitations of the current study are:

- The ROIs were manually selected, so the current CAD system is not fully automatic. In future work on breast tissue classification we will use an automatic approach to select the suspicious ROIs from mammograms. In the case of breast density classification, we will use ULDP to segment dense regions from the breast area, and then the breast density will be calculated as a percentage between the area of the dense tissue and the area of the whole breast.
- We have evaluated the proposed descriptor with ROIs extracted from the middle of the breast area; however, in some cases the masses are located close to the pectoral muscle. Thus, the extracted ROIs may contain a small part from the pectoral muscle which is white in mammograms (like dense tissues). In these cases, we are not sure whether the proposed descriptor will give good mass/normal classification results or not.
- The proposed descriptor was evaluated with mass/normal breast tissue classification in breast ultrasound images. The result was not good because of the presence of speckle noise. In chapter 6 we propose the fuzzy local directional pattern to improve the results of mass detection in ultrasound images.

4.6 Conclusion

In this chapter we proposed a CAD system for analyzing breast tissues in mammograms. The analysis included breast tissue classification and breast density classification. We proposed the ULDP descriptor for feature extraction. This descriptor encodes a neighborhood in the breast region on the basis of its edge responses and spatial information. The ULDP histogram can characterize breast masses as well as different tissues in the breast. Two main experiments have been performed to evaluate ULDP with two databases: mass/normal breast tissue classification and breast tissue density classification using mini-MIAS (film-screen

database) and INbreast (full field digital database). Our experiments enable us to draw the following conclusions:

- ULDP can properly discriminate between different tissues in mammograms regardless of their *sizes*, *shapes* or *margins*. Results were good for both mass/normal breast tissue classification and breast tissue density classification. In addition, breast tissue density classification based on the ULDP histogram yielded a high Kappa coefficient.
- The correct classification rate based on the ULDP descriptor with digital mammograms was higher than the rate with film-screen mammograms, due to the good contrast of digital mammograms.
- We studied the effect of breast density on the performance of the proposed descriptor. ULDP performed more or less constantly with different breast densities.

In the next chapter we study the impact of such factors as pixel resolution, integration scale, preprocessing and feature normalization on the performance of texture methods for breast tumor classification.

CHAPTER 5

The impact of image resolution, patch size, preprocessing and feature normalization on texture analysis for tumor classification

5.1 Introduction

CAD systems are typically used to analyze mammograms in screening. While radiologists are generally pleased with the performance of CAD for clustered microcalcification detection, they have little confidence in CAD for mass detection. The most common complaint by radiologists is that CAD systems lead to a large number of false positives (Gilbert et al., 2008).

Although mammography is a highly sensitive method for the early detection of

breast cancer, specificity is low in the classification of benign and malignant masses. Texture analysis methods are one of the options for improving the specificity of classification algorithms applied to mammography. These methods may provide additional information in distinguishing benign and malignant masses. Although several feature extraction methods have been proposed for analyzing mammograms, improving the classification performance remains a challenge.

Texture analysis methods have been widely used to analyze mammographic images because they produce information about the spatial arrangement of intensities in the mammogram. Texture is one of the major mammographic characteristics for mass classification. For instance, several studies have used texture analysis methods to distinguish between normal and abnormal tissue (Bellotti1 et al., 2006; Melendez et al., 2014; Oliver et al., 2007; Pomponiu et al., 2014; Zheng, 2010), or to discriminate between benign and malignant masses (Chan et al., 1997; Rangayyan et al., 2010; Soltanian-Zadeh et al., 2004). Other studies have used texture analysis methods to estimate breast density (Oliver et al., 2008) or to segment masses from mammograms (Oliver et al., 2010).

CAD systems usually focus on a ROI to study breast masses. The texture of this ROI describes the pattern of spatial variation of grey levels in a neighborhood that is smaller than the breast area but big enough to include the masses. In other words, texture must be analyzed in a region, and the size of this region should be tuned. Thus, the following question needs to be answered: *what is the optimal neighborhood size (integration scale) for texture analysis*? In addition, the size of a mammogram is usually in the range of thousands of pixels. Consequently, several works have reduced the original resolution of a mammogram to reduce the computational complexity and the execution time of their algorithms (Sampat et al., 2008), or to save resources (e.g. memory and storage space). However, image downsampling may also affect the performance of the texture analysis methods. Therefore, the following question also needs to be answered: *how far can we downsample the image without affecting the performance of the texture methods*?

In breast cancer CAD systems, such preprocessing operations as image filtering

or enhancement are usually applied to mammograms. Pisano et al. (1998) show that the contrast-limited adaptive histogram equalization (CLAHE) applied to a mammogram before it is displayed can make the indicative structures of breast cancer more visible. Sharpening (SH) is used to improve the detection of clustered calcifications (Anand et al., 2013). The median filter (MF) is used to remove the noise from the mammograms (Subashini et al., 2010). *Preprocessing may affect the performance of texture analysis methods* because it effectively changes the gray levels of the images. This effect should be assessed. After extracting the texture features from a given mammogram, they are usually normalized before proceeding to the classification stage. *The normalization method used may also affect the final classification results.*

In this chapter, we study the effect of pixel resolution, integration scale, preprocessing and feature normalization on the performance of texture analysis methods when used to classify masses in mammograms. To this end, we have chosen five widely/recently used texture methods: LBP, LDN, HOG, Haralick's features (HAR) and Gabor filters (GF). In order to evaluate the performance of these methods, we extracted a set of ROIs containing lesions from the mini-MIAS database (Suckling et al., 1994), and we used each texture analysis method to classify the ROIs into benign or malignant. The performance of each texture method is evaluated with five pixel resolutions (200 μm, 400 μm, 600 μm, 800 μm, 1000 μm), six integration scales (25×25, 32×32, 50×50, 64×64, 75×75, 100×100 pixels), three preprocessing steps (CLAHE, MF, SH), and five feature normalization methods. Linear and nonlinear SVM classifiers are also used.

To the best of our knowledge, only one previous study has conducted a similar evaluation. Rangayyan et al. studied the effect of pixel resolution on texture features of breast masses in mammograms (Rangayyan et al., 2010). However, they only took pixel resolution and Haralick's features into account. In contrast, the current study takes into account a wider range of factors–for example, pixel resolution, integration scale, preprocessing and feature normalization–and considers a larger number of more powerful texture descriptors that have been successfully applied in recent work. We

also include linear and nonlinear SVMs so that both relatively simple and complex classification approaches can be assessed. Lastly, we analyze the combination of the best options for those factors using three approaches: the greedy method, sequential forward selection (SFS) and exhaustive search (ExS).

5.2 Methods

In this study, we assess the performance of five texture analysis methods (LBP, LDN, HOG, HAR, GF) while varying the pixel resolution, integration scale, image preprocessing algorithm and data normalization method. To this end, we extracted a set of ROIs containing either benign or malignant masses from the mini-MIAS database. Given a particular texture analysis method, a feature vector is extracted from each ROI and fed into a linear support vector machine (LSVM) or a nonlinear support vector machine (NLSVM). The trained models are used to determine if an unseen ROI contains a benign or a malignant mass.

The mini-MIAS database, consisting of 322 mediolateral oblique images of 161 cases, is used in our experiments. It was created from the original MIAS database by downsampling the images from 50 μm to 200 μm per pixel and clipping/padding to a fixed size of 1024×1024 pixels. A ground truth was prepared by experienced radiologists and confirmed using a biopsy procedure. In this study 109 ROIs, 60 containing a benign mass and 49 containing a malignant mass, were used. Fig. 5.1 shows examples of the extracted ROIs.

(a) **(b)**

Figure 5.1: ROIs extracted from the mini-MIAS breast cancer database. A ROI containing: (a) a benign mass and (b) a malignant mass

The authors of the mini-MIAS database reported that they reduced the pixel

resolution of the original MIAS database (digitized at 50 μm) to 200 μm by popular request. Moreover, several studies have used the pixel resolution 200 μm as a baseline resolution in their applications (Karssemeijer, 1998; Sampat et al., 2008). We do the same in here.

5.2.1 Texture analysis methods

This section explains the texture analysis methods used and the parameters selected for each of them.

- **Local binary pattern**. In this study, a 3×3 neighborhood is used to generate the histogram of uniform LBPs for each ROI. The uniform mapping produces 59 output labels (59 dimensions) for neighborhoods of 8 pixels.
- **Local directional number.** In the LDN (Ramirez Rivera et al., 2013), the edge responses are computed in eight different directions by convoluting the Kirsch compass masks Kirsch (1971) with the ROIs. The location of the top positive and negative edge responses are used to generate a 6-bit code for each pixel. Finally, the histogram of the LDN codes is calculated in the given ROI (64 dimensions).
- **Histogram of oriented gradients.** In order to get the best performance out of HOG, its parameters have been empirically tuned. In this study, we used a 3×3 cell size, 8×8 cells for the block size, and a 9-bit histogram.
- **HAR features.** A GLCM is computed from each ROI, and then 14 texture features are calculated: *angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measure of correlation 1, information measure of correlation 2* and *maximal correlation coefficient* (Rangayyan et al., 2010). The mathematical expression of each feature can be found in Section 2.3.2.3.
- **Gabor filters.** In this study, we used 4 scales and 6 orientations to obtain these filtered ROIs. This design produces 24 responses. For each ROI, the energies of the 24 responses are calculated, and then they are aggregated in

order to form the feature vector.

5.2.2 Preprocessing

The performance of the texture analysis methods is evaluated with three preprocessing algorithms: CLAHE, median filter (MF) and sharpening (SH).

- *CLAHE*: it works on small regions of the input ROI (known as tiles). The contrast of each tile is enhanced, so the histogram of the output region approximately matches a pre-defined distribution (Puff et al., 1994). In this study, the *Rayleigh* distribution is used Pisano et al. (1998).
- *MF*: each pixel in the filtered ROI contains the median value of the $m \times n$ neighborhood around the corresponding pixel in the input ROI Subashini et al. (2010). In this study, a 3×3 neighborhood is used.
- *SH*: in order to sharpen a ROI, it is first blurred; then, the edges are detected in the blurred ROI and added to it to produce a sharper image (Anand et al., 2013).

Fig. 5.2 shows examples for MF, SH and CLAHE when they are applied to benign and malignant masses.

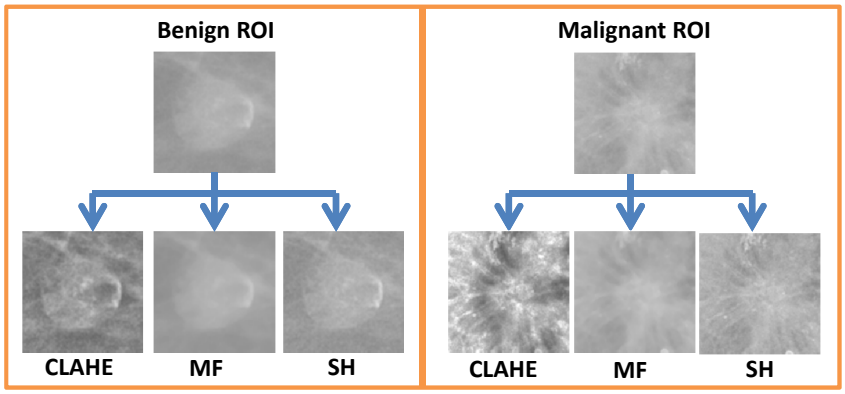


Figure 5.2: Examples of ROI preprocessing

5.2.3 Feature normalization methods

Feature vectors are normalized in order to prevent attributes with higher numeric ranges from dominating those with lower numeric ranges. Given a feature vector

$x = [x_1, x_2, x_3, \ldots, x_N]$, the normalized feature vector x_{new} is calculated using five normalization methods as follows (Aksoy and Haralick, 2001; Juszczak et al., 2002):

- The zero mean unit variance (zs) method: $x_{new} = (x - \mu)/\sigma$, where μ and σ are the mean and the variance of x.
- The maximum-minimum (mn) method: $x_{new} = (x - x_{min})/(x_{max} - x_{min})$, where x_{max} and x_{min} are the maximum and minimum of x.
- The $l1$ method scales x to unit length using the ℓ^1-norm, $x_{new} = x/\sum_{n=1}^{N} |x_n|$.
- The $l2$ method scales x to unit length using the ℓ^2-norm, $x_{new} = x/\sqrt{\sum_{n=1}^{N} |x_n|^2}$.
- The nh method scales x to unit length as follows, $x_{new} = x/\sum_{n=1}^{N} x$.

Fig. 5.3 shows examples of normalizing LBP features extracted from a ROI that contains a benign mass using zs, mn, $l1$, $l2$, nh normalization methods.

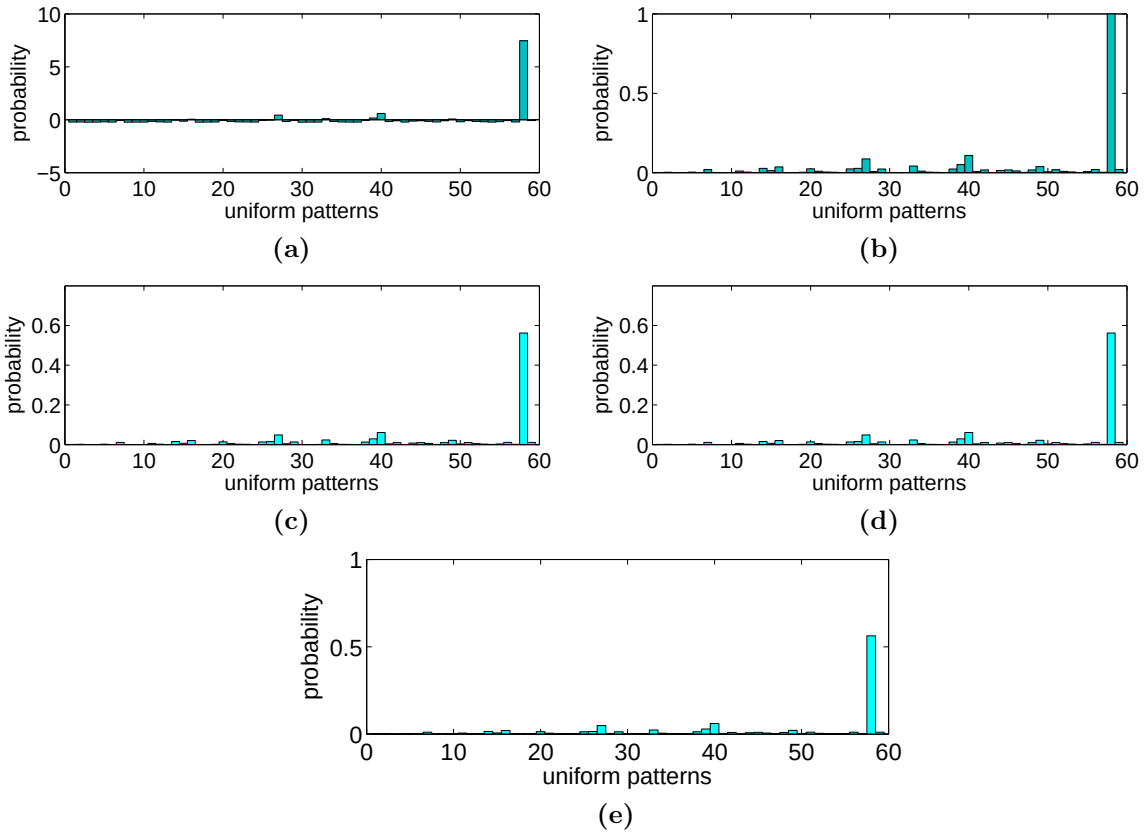


Figure 5.3: Feature normalization using (a) zs, (b) mn, (c) $l1$, (d) $l2$ and (e) nh normalization methods

5.3 Experimental results

In this section, we present the effect of pixel resolution, integration scale, preprocessing steps, and normalization methods on the performance of the texture analysis methods when they are applied to *benign/malignant* mass classification in mammograms. We also study the effect of different combinations of the aforementioned factors.

The performance of each texture analysis method is measured in terms of AUC (Fawcett, 2006). The SVM classifier provides decision values related to the membership of each class. To generate a ROC curve, we vary a threshold over the decision values. We also use the k-fold cross validation technique to generate the training and testing data. In this study, k=10. The mean AUC value is calculated over the cross validation process.

5.3.1 Effect of pixel resolution and integration scale

As we commented in Section 5.2, the pixel resolution 200 μm has been widely used in several studies (Karssemeijer, 1998; Sampat et al., 2008). So, in this experiment we start with this pixel resolution and then the mammograms are downsampled to generate different pixel resolutions. The downsampling step includes anti-aliasing filtering and a bicubic interpolation. Five pixel resolutions are generated (200 μm, 400 μm, 600 μm, 800 μm, 1000 μm), and then we use six integration scales (25×25, 32×32, 50×50, 64×64, 75×75, 100×100 pixels) to analyze the texture of each ROI. In this experiment, no preprocessing is applied, and the standard *zs* normalization method is used to normalize the extracted feature vectors. The effect of pixel resolution and integration scale on the performance of LBP, LDN, HOG, HAR and GF with the LSVM and the NLSVM is shown in Fig. 5.4.

As shown in Fig. 5.4, each texture method gives its best AUC value at a particular pixel resolution and integration scale. Of all the texture methods, LBP gives the best AUC value (0.78) at pixel resolution 800 μm and integration scale 75×75.

The analysis of variance (ANOVA) test (Armitage et al., 2008) has been used

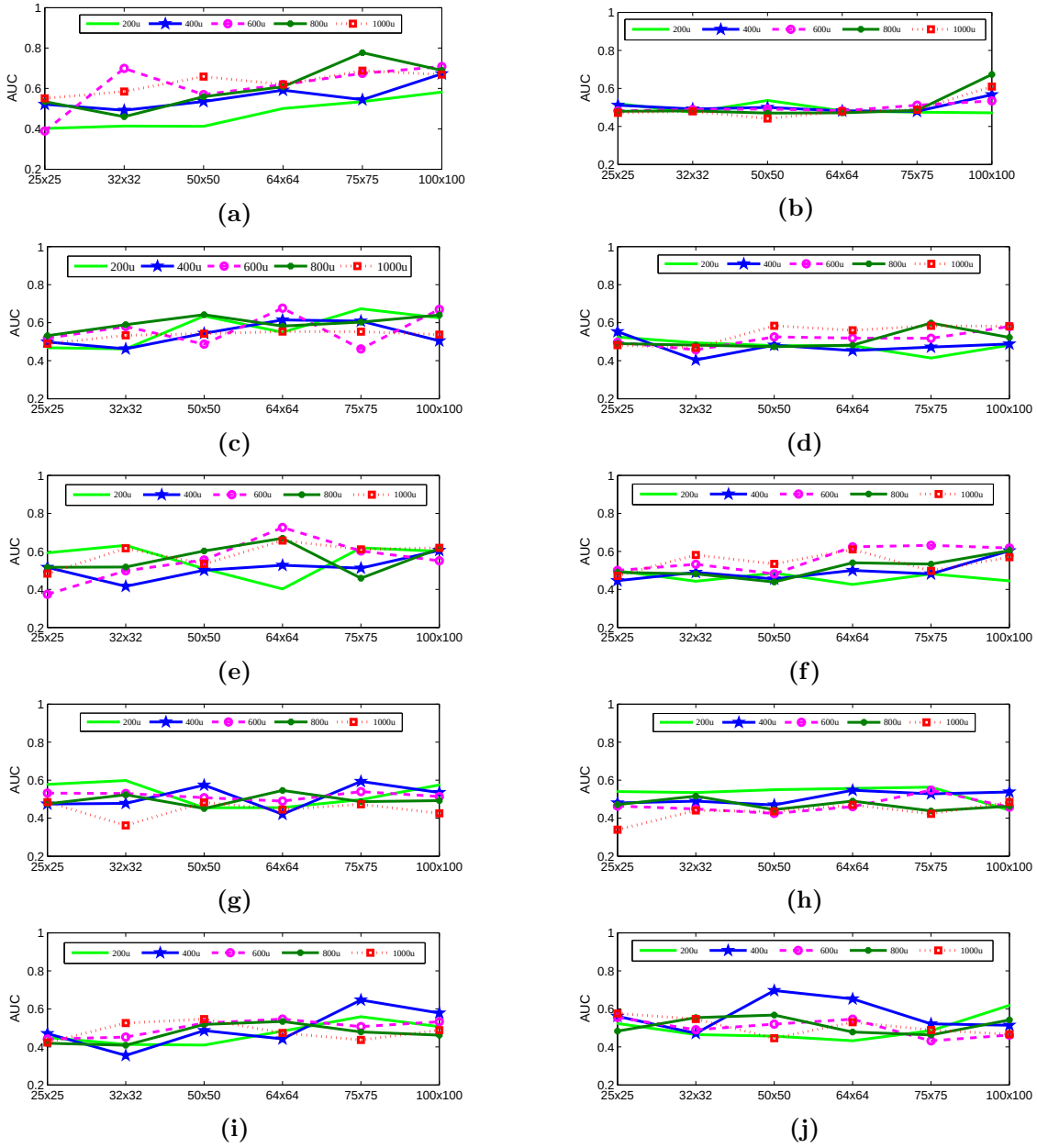


Figure 5.4: The effect of pixel resolution and integration scale on the performance of the texture methods with LSVM (left) and NLSVM (right), (a)-(b) LBP, (c)-(d) LDN, (e)-(f) HOG, (g)-(h) HAR and (i)-(j) GF

to examine the interaction between pixel resolutions and integration scales. The experimental design of ANOVA includes two factors: pixel resolution (Res) and integration scale (IS). Res has five levels (200 μm, 400 μm, 600 μm, 800 μm, 1000 μm), whereas IS has six (25×25, 32×32, 50×50, 64×64, 75×75, 100×100 pixels). Each combination of the levels of Res and IS produces an AUC value (response). The confidence level is set to 0.05. The results are shown in Tables 5.1 and 5.2.

Table 5.1: Summary of the ANOVA results of pixel resolution and integration scale with the LSVM (the value in each cell is a p-value)

Method	Res	IS	Res*IS
LBP	0.0024	0.001	0.908
LDN	0.1174	0.4035	0.8037
HOG	0.3905	0.6515	0.4636
HAR	0.7846	0.0962	0.2895
GF	0.083	0.8259	0.9864

Table 5.2: Summary of the ANOVA results of pixel resolution and integration scale with the NLSVM (the value in each cell is a p-value)

Method	Res	IS	Res*IS
LBP	0.9332	0.0101	0.0095
LDN	0.2387	0.0772	0.6451
HOG	0.0448	0.0103	0.5138
HAR	0.4253	0.004	0.0847
GF	0.6552	0.3109	0.2024

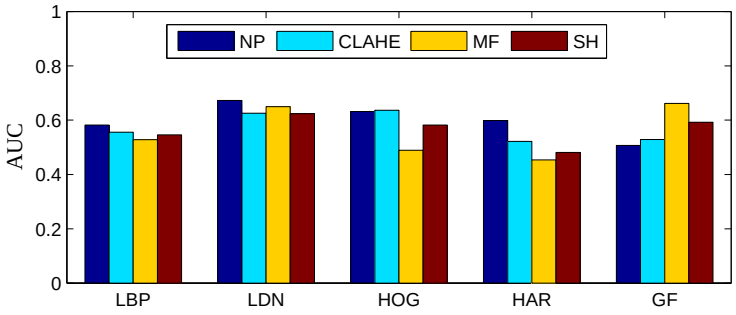
As shown in Table 5.1, with LBP and the LSVM, the mean responses for the levels of pixel resolution are significantly different (p=0.0024). Similarly, the mean responses for the integration scale levels are significantly different. In the case of LDN, HOG, HAR and GF, the mean responses for the pixel resolution and integration scale levels are not significantly different. The p-values indicate that the interaction between the levels of pixel resolution and integration scale (Res*IS) are not significant.

As shown in Table 5.2, the mean responses for the levels of pixel resolution are significantly different in the case of HOG with NLSVM. In the case of LBP, LDN, HAR and GF, the mean responses for the levels of pixel resolution are not significantly different. The mean responses for the integration scale levels are significantly different in the case of LBP, HOG and HAR. With LBP and NLSVM,

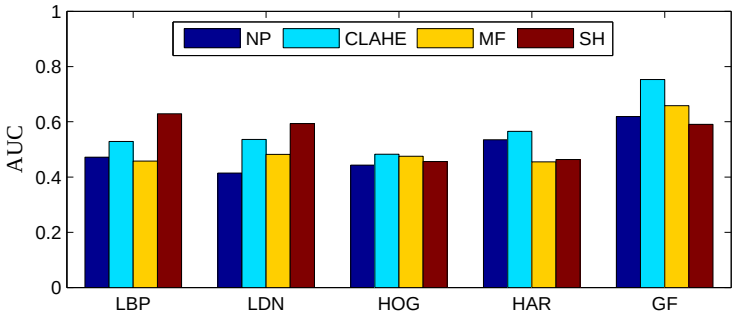
the interaction between pixel resolution and integration scale (Res*IS) is significant.

5.3.2 Effect of preprocessing

In this experiment, the integration scale that gave the highest AUC value with each texture analysis method at the baseline pixel resolution of 200 μm and the standard *zs* normalization method are used. The effect of no preprocessing (NP), CLAHE, MF and SH on the performance of each texture analysis method is shown in Fig. 5.5. As can be seen, each texture method produces the highest AUC value with a particular preprocessing algorithm. In this experiment, LBP gives the highest AUC value with SH and NLSVM, while LDN and HAR give the highest AUC value with NP and LSVM. HOG gives the highest AUC value with CLAHE and LSVM. In turn, GF gives the highest AUC value with CLAHE and NLSVM.



(a)



(b)

Figure 5.5: Texture analysis methods with NP, CLAHE, MF and SH using (a) LSVM, and (b) NLSVM

5.3.3 Effect of feature normalization methods

In this experiment, we study the effect of five normalization methods (zs, mn, $l1$, $l2$, nh) on the performance of each texture analysis method. For each texture analysis method, we use the integration scale that gives the highest AUC value at pixel resolution 200 μm. No preprocessing method is used. The effect of the normalization methods is shown in Fig. 5.6. With the LSVM, zs normalization has led LBP and LDN to AUC values that are better than other normalization methods, while GF gives its highest AUC value with $l1$ normalization and NLSVM. As shown in the figure, each texture analysis method gives its highest AUC value with a particular normalization method.

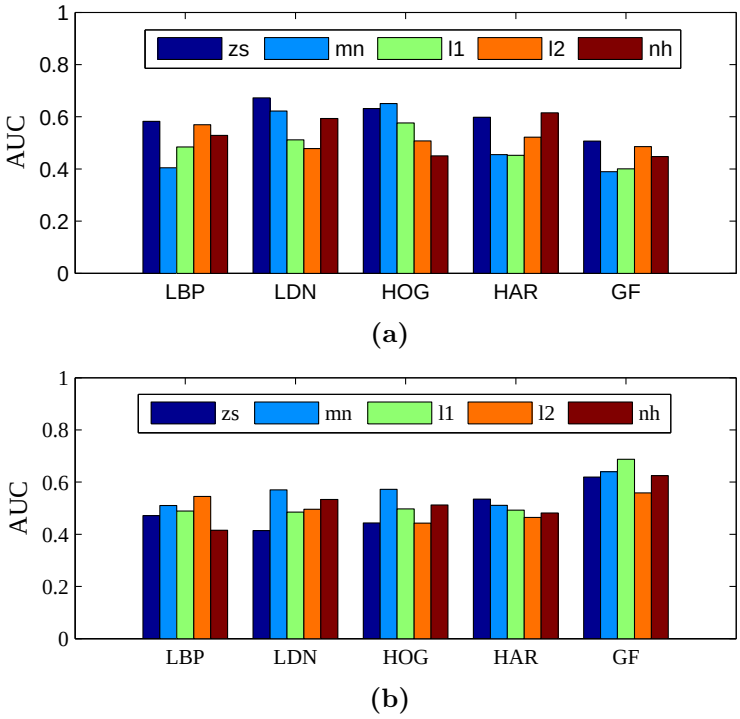


Figure 5.6: Texture analysis methods with different feature normalization methods using (a) LSVM, and (b) NLSVM

5.3.4 Summary of the results

The best AUC values of each texture analysis method for the experiments in Sections 5.3.1, 5.3.2 and 5.3.3 are summarized in Table 5.3. LBP produces the

best AUC value (0.78) at pixel resolution 800 μm, integration scale 75×75, no preprocessing, zs normalization method and LSVM. In turn, HAR produces the lowest AUC value (0.61). LBP, LDN, HOG and HAR give their best values with LSVM, whereas GF gives its best AUC value with NLSVM.

Table 5.3: Best AUC value for each texture analysis method and the configuration that yields it considering the experiments in Sections 5.3.1, 5.3.2 and 5.3.3

Method	Best value	Res(μm)	IS	classifier	preprocessing	Normalization
LBP	0.78	800	75×75	$LSVM$	NP	zs
LDN	0.68	600	64×64	$LSVM$	NP	zs
HOG	0.72	600	64×64	$LSVM$	NP	zs
HAR	0.61	200	32×32	$LSVM$	NP	nh
GF	0.75	200	100×100	$NLSVM$	$CLAHE$	zs

5.3.5 Combining the levels of all factors

To find the best combination among the levels of all factors, we use three approaches: the greedy approach, sequential forward selection (SFS) and exhaustive search (ExS). In the greedy approach, we try to combine the best options of the aforementioned factors. For each texture analysis method, we summarize the best levels of pixel resolution, integration scale and normalization methods in Table 5.4.

Table 5.4: The best option of pixel resolution, integration scale, preprocessing and normalization methods with each texture method

Method	Res (μm)	IS (pixels)	Preprocessing	Normalization
LBP	800	75×75	SH	zs
LDN	600	64×64	NP	zs
HOG	600	64×64	CLAHE	mn
HAR	200	32×32	NP	nh
GF	400	50×50	CLAHE	$l1$

Table 5.5 shows that combining the best levels of pixel resolution, integration scale, preprocessing and feature normalization does not improve the AUC values of the texture analysis methods reported in Table 5.3. In fact, LBP, HOG and GF produced substantially lower AUC values. LSVM yields higher AUC values than NLSVM.

Secondly, we use an SFS approach to find the best combination. This consists of two sequential steps: finding the normalization method that most improves

Table 5.5: Results of the greedy approach (AUC)

Method	LSVM	NLSVM
LBP	0.46	0.40
LDN	0.68	0.52
HOG	0.44	0.44
HAR	0.61	0.48
GF	0.58	0.54

the current performance, and then finding the preprocessing method that keeps improving this performance. For each texture method, in the first step, we start with the best pixel resolution and integration scale summarized in Table 5.4. Then, with no preprocessing, the extracted features are separately normalized by each normalization method. Then, the one that improves the performance in combination with the previous two factors is added. In the second step, we apply each preprocessing option to the ROIs (NP, CLAHE, MF, SH). Then we extract the texture features and normalize them using the best normalization method obtained in the previous step. Both LSVM and NLSVM are used to classify the ROIs. Table 5.6 shows that the SFS does not improve the AUC value of GF achieved in Table 5.3. LBP, LDN, HOG and HAR give AUC values close to the ones listed in Table 5.3. With all the texture methods, the SFS approach gives better AUC values than the greedy approach.

Table 5.6: Results of the SFS approach

Method	Best AUC	Best parameters
LBP	0.780	zs, NP, LSVM
LDN	0.679	zs, NP, LSVM
HOG	0.716	zs, NP, LSVM
HAR	0.605	nh, NP, LSVM
GF	0.720	zs, CLAHE, NLSVM

Lastly, we use an ExS algorithm, which searches for the best combination of five pixel resolutions, six integration scales, four preprocessing methods (NP, CLAHE, MF and SH) and five data normalization methods, and finds a total of 600 combinations. In the previous experiments, we found that LSVM usually gives the

best results except with GF. NLSVM has two parameters that need to be optimized to provide the best classification results. Adding NLSVM's parameter optimization to the ExS substantially increases its complexity. So we decided to use only LSVM in this final test. As shown in Table 5.7, the ExS approach improves the AUC values of LDN, HOG and HAR. The GF gives an AUC value lower than the one listed in Table 5.3 because the LSVM can not perfectly separate the GF features.

Table 5.7: Results of the ExS approach

Method	Best AUC	Best parameters
LBP	0.78	800, 75×75, NP, zs
LDN	0.70	600, 75×75, MF, zs
HOG	0.737	1000, 50×50, SH, mn
HAR	0.666	800, 32×32, CLAHE, nh
GF	0.691	600, 32×32, NP, *l1*

5.4 Discussion

The performance of texture analysis methods when applied to benign/malignant mass classification can be affected by many factors. In this study, we focus on the effect of factors such as pixel resolution, integration scale, preprocessing and feature normalization. We use the well-known mini-MIAS database. We start with the original pixel resolution of the mini-MIAS database (200 μm), and then we downsample the mammograms in order to generate the pixel resolutions 400 μm, 600 μm, 800 μm, 1000 μm. In addition, six integration scales are used (25×25, 32×32, 50×50, 64×64, 75×75, 100×100 pixels). These integration scales cover most of the sizes of the masses in the mini-MIAS database, which range from a few pixels to tens of pixels (the mean diameter of the circle containing the masses is about 49 pixels). Several previous studies have used one of these integration scales to analyze the texture of mammograms (Oliver et al., 2007; Pomponiu et al., 2014). Thus, we hypothesize that the aforementioned integration scales are able to deal with all the masses appearing in the mini-MIAS database.

The shape of breast masses is one of the powerful features that can be used to

discriminate between benign and malignant masses. The boundaries of malignant masses usually have irregular shapes, while the boundaries of benign masses have regular ones. In the case of breast mass analysis, pixel resolution may be a critical factor because image downsampling may remove some fine detail from the image. However, as our results indicate, it would be possible to decrease the resolution far beyond 200 μm and still obtain good classification results. A notable example is LBP, which actually performed best at 800 μm. One possible explanation for this is that core information such as that contained in the boundary of masses may still be preserved even after downsampling and be more useful for methods such as LBP that operate over higher order statistics of grey intensity values. Obviously, when the resolution is far too low, the classification performance degrades, as the shape of the boundaries of benign and malignant masses will be very similar. Another important factor is the integration scale, as it should be big enough to cover the masses and their boundaries, and small enough to exclude other tissues. The effect of pixel resolution and integration scale on the performance of texture methods should be studied jointly.

As summarized in Table 5.4, each texture method gives its highest AUC value at a particular pixel resolution and integration scale. HAR gave its highest AUC value at a pixel resolution of 200 μm and an integration scale of 32×32 pixels. In turn, LBP gave its best AUC value at a pixel resolution of 800 μm and an integration scale of 75×75 pixels. The integration scale and pixel resolution interact with each other in a particular way. In the case of LBP, LDN and HOG, the texture features of each method are represented in a histogram. This histogram contains the repetition of the patterns detected by each method at a particular pixel resolution and integration scale. LBP features calculated at a pixel resolution of 200 μm are different from those calculated at a pixel resolution of 400 μm. LDN and HOG also produce different patterns at different pixel resolutions. The local patterns of LBP, LDN, HOG are usually calculated within a particular integration scale. Different integration scales will yield different histograms for the local patterns. For instance, the histograms of LBP that are calculated with the integration scales 75×75 and 100×100 are

different.

ANOVA results show that the mean AUC values of the pixel resolutions are significantly different in the case of LBP with LSVM. In addition, the mean AUC values of the integration scales are significantly different with LBP, HOG and HAR, and NLSVM. The performance differences with respect to the pixel resolutions and the integration scales are only significantly different with the LBP and the NLSVM (p=0.0095). These results indicate that the choice of pixel resolution and integration scale has a direct effect on the performance of a texture-based CAD system, because of its effect on the texture method used.

Image preprocessing also affects the performance of the texture analysis methods. HOG and GF give the highest AUC values with CLAHE, while LDN and HAR perform better with NP. Indeed, CLAHE, MF and SH change the intensities of the mammograms in different ways. As a result, each texture analysis method will produce a different AUC value with each preprocessing technique. In general, the preprocessing approach that makes the small-scale structures in the ROIs more visible would give the texture methods more discriminative power. For instance, CLAHE leads GF to its best AUC value (0.75). There is also a relation between the principle of operation of some texture methods and the preprocessing used. For instance, the binary patterns of the LDN are calculated on the basis of the edge responses of each pixel in the image. MF removes the outliers before calculating the edge responses. Thus, the edge responses will be properly calculated, and the discriminative power of LDN will improve.

Prior to mass classification, the calculated texture features should be normalized to prevent attributes with higher numeric ranges from dominating those with lower numeric ranges. As shown in our experiments, each texture method produces its highest AUC value with a particular normalization method. This is because each normalization method produces numerical values with different distributions (see Fig. 5.3). Consequently, the arrangement of the texture features in the feature space with one normalization method is different than with other normalization methods. Thus, the normalization method changes the final values of the features computed

by each texture method. As shown in Table 5.4, LBP and LDN give the highest AUC values with zs normalization, HOG with mn, HAR with nh, and GF with $l1$. In the classification stage, we use two widely used classifiers in the field of mammogram analysis: LSVM and NLSVM. The first one linearly separates the texture features in the feature space, while the second one uses a kernel function (RBF). As shown in Table 5.3, LBP, LDN, HOG and HAR give their highest AUC values with LSVM. GF, however, gives its best AUC value with NLSVM, indicating that GF features are not linearly separable.

Table 5.3 shows a summary of the levels of pixel resolution, integration scale, preprocessing and normalization methods that have led each texture method to its best AUC value in the experiments in Sections 5.3.1, 5.3.2 and 5.3.3 . HAR and GF give the best AUC values at a pixel resolution of 200 μm, while LDN and HOG give their best results at a pixel resolution of 600 μm. No method gives its best AUC value with the integration scales 25×25 and 50×50 pixels.

The greedy, SFS and ExS approaches are used to find the best combination of the levels of all factors. Although the greedy approach is the least complex approach, it yielded poor AUC values. In contrast, the ExS gave good results, but computationally it is the most complex. The SFS approach provides a trade-off between accuracy and computational complexity. It is not as complex as the ExS approach and it does not produce such poor AUC values as the greedy approach. In the case of LBP, LDN, HOG, HAR, Table 5.6 shows that the SFS approach produces approximately the same results as those obtained with the ExS approach. The GF gave better AUC values with the SFS approach because it used NLSVM, whereas with the ExS approach the calculation of the optimal values of its internal parameters was a challenge.

Rangayyan et al. extracted 111 ROIs from mammograms obtained from three different sources: the mammographic image analysis society (MIAS), the teaching library of the Foothills Hospital in Calgary, and a screening test (the Alberta program for the early detection of breast cancer) (Rangayyan et al., 2010). Although using mammograms from different sources may be helpful to assess the robustness of the

texture methods studied, the three mammogram sets used by Rangayyan et al. were digitized at different pixel resolutions. Thus, the characteristics of the textures extracted from the 111 ROIs may be different. This changes the characteristics of the features extracted, so the effect of pixel resolution on the performance of the texture methods may have not been properly studied. In contrast, in the current study, the ROIs were extracted from a single source (the mini-MIAS database). Rangayyan et al. extracted ROIs of different sizes (each ROI included a mass) and they did not mention the effect of the integration scale on the performance of the texture methods. Conversely, the current study has considered six integration scales. With a pixel resolution of 800 μm, an integration scale of 75×75, no preprocessing, the zs normalization method and LSVM, the LBP gives a better AUC value (0.78) than other texture methods, exceeding the best AUC value (0.75) achieved by Rangayyan et al. (2010).

5.5 Conclusion

When applied to benign/malignant mass classification in mammograms, texture analysis methods are sensitive to changes in pixel resolution, integration scale, preprocessing and feature normalization. The best combination of these factors should be identified so as to obtain the best discriminative power for each texture analysis method. We expect that the assessment performed in this study will help researchers to do this. Because of its computational cost advantage, sequential forward selection would be a suitable approach for determining a reasonable (possibly the best) factor configuration.

In the next section we propose the fuzzy local directional pattern for analyzing breast cancer in ultrasound images. We also propose the use of super-resolution approaches to improve the performance of texture methods when applied to tumor classification in ultrasound images.

Part III

Analysis of breast cancer in ultrasound images

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

CHAPTER 6

Breast tissue characterization in ultrasound images using fuzzy local directional patterns

6.1 Introduction

In a mammography, each breast is compressed using compression plates, and then it is X-rayed from top to bottom or at an angle. Sonographies are safer and painless, and they generate real time images of the inside of the breast using ultrasound waves. *Breast density* is one of the main failure factors of mammographies because dense tissues may hide some tumor regions. Breast density represents the relative amounts of fibroglandular and fat tissue in a woman breast (Lokate et al., 2010). The

well-known mini-MIAS breast cancer database (Suckling et al., 1994) classifies breast density into three categories: fatty, fatty-glandular and dense glandular. *Dense* breasts have more glandular and fibrous tissues, and they show up white in the mammogram. Therefore, they hide cancer regions, which also usually show up white in mammograms. In turn, *fatty* breasts have more fatty tissues and they show up grey in mammograms. Thus, it is easy to detect cancer in fatty breasts. Indeed, sonographies are better than mammographies at detecting abnormalities in dense breasts. Sonographies have provided important support to mammographies in breast cancer detection. They cannot replace a mammogram for breast screening, but they can provide physicians with more help.

Several CAD systems have been proposed to analyze breast cancer in ultrasound and X-ray images. Shi et al. (2010) discuss the approaches used in ultrasound breast images CAD stages and summarize their advantages and disadvantages. A breast cancer CAD system based on a fuzzy support vector machine is developed in (Shi et al., 2010) to automatically detect masses in ultrasound images. Moreover, *fuzzy local binary patterns (FLBP)* are proposed in (Keramidas et al., 2011). FLBP incorporate fuzzy logic in the representation of local patterns in ultrasound images. FLBP are extracted from a set of ROIs acquired from thyroid ultrasound images, and then they are classified by a SVM classifier into nodule or non-nodule classes.

Noise, breast density and variation in breast compressions give the breast tissues a fuzzy appearance. Unfortunately, the literature shows no consensus on an optimal feature set for mass/normal breast tissue classification, which means that the methods proposed in the literature do not produce a complete characterization of different tissues in breast images, so the number of false positives is high.

In this chapter, we propose the fuzzy local directional pattern (FLDP) for characterizing breast tissues. FLDP is an extension of the ULDP (chapter 4). The rationale behind the use of fuzzy logic is to compensate for the uncertainty of the visual appearance of breast tissues due to noise, breast density and variation in breast compressions. FLDP describes the shapes, margins, spots, edges, corners, junctions and other structures of different tissues in a breast region. FLDP is

evaluated with mass/normal classification of ROIs extracted from ultrasound images and mammograms.

6.2 Fuzzy local directional pattern

The FLDP describes a given pixel through its edge responses. The edge responses ER of each pixel are computed using the Kirsch compass masks (see Fig.6.2). Given a particular pixel, its *eight* edge responses ER can be defined as:

$$ER = \{ER_0, ER_1, \dots ER_7\} \in \mathbb{R}^8 \tag{6.1}$$

Let A be the set of ER greater than zero, and B the set of ER smaller than zero:

$$A \equiv \{ER_i \mid 0 \leq i \leq 7,\ \ ER_i \geq 0\} \tag{6.2}$$

$$B \equiv ER - A \equiv \{ER_i \mid 0 \leq i \leq 7,\ \ ER_i < 0\} \tag{6.3}$$

In crisp approaches, a hard threshold is used to determine the prediction of each variable. In turn, fuzzy logic allows the use of a membership function to determine the class of each variable. In the fuzzification process, each input variable is mapped to its corresponding fuzzy variable according to a set of fuzzy rules (Zadeh, 1965). A and B can be defined as two fuzzy sets $\widetilde{A}$ and $\widetilde{B}$, where $\widetilde{A}$ contains the *positive* edge responses, and $\widetilde{B}$ contains the *negative* edge responses. The fuzzy sets $\widetilde{A}$ and $\widetilde{B}$ can be expressed as:

$$\widetilde{A} \equiv \{(ER_i, \mu_+(ER_i)) \mid 0 \leq i \leq 7,\ \ \mu_+(ER_i) \geq \mu_-(ER_i)\} \tag{6.4}$$

$$\widetilde{B} \equiv \{(ER_i, \mu_-(ER_i)) \mid 0 \leq i \leq 7,\ \ \mu_-(ER_i) > \mu_+(ER_i)\} \tag{6.5}$$

Given an edge response, its degree of membership to these fuzzy sets can be computed. We used a linear function to calculate the degree of each edge response ER_i to be *negative*, or the degree of each ER_i to be *positive* (see Fig. 6.1). Let μ_+

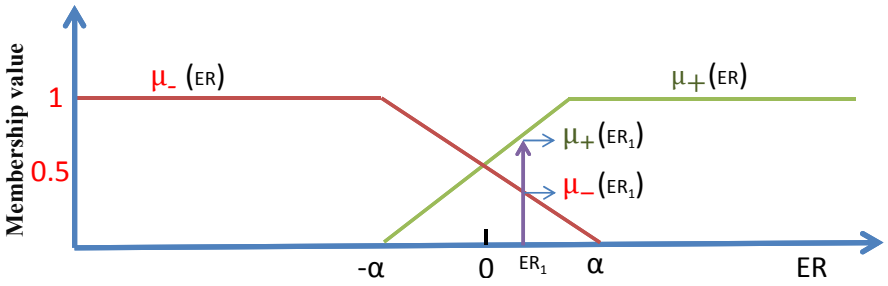


Figure 6.1: Linear membership functions, the green curve represents the membership function of the *positive* fuzzy set, whereas the blue curve represents the membership function of the *negative* fuzzy set

define the degree of each ER_i to be *positive*:

$$\mu_+(ER_i) = \begin{cases} 0 & \text{if } ER_i < -\alpha \\ 0.5 + ER_i/2\alpha & \text{if } -\alpha \leq ER_i \leq \alpha \\ 1 & \text{if } ER_i > \alpha, \end{cases} \tag{6.6}$$

where α is a threshold. In addition, μ_- defines the degree of each ER to be *negative*:

$$\mu_-(ER_i) = 1 - \mu_+(ER_i) \tag{6.7}$$

Given the vector ER of a certain pixel, the degree of membership of each edge response to the *positive* and *negative* fuzzy sets can be computed. Each edge response ER_i may belong to one of the following categories:

a) $\mu_+(ER_i) = 1, \;\; \mu_-(ER_i) = 0, \;\; [ER_i \geq \alpha]$

b) $\mu_+(ER_i) = 0, \;\; \mu_-(ER_i) = 1, \;\; [ER_i \leq -\alpha]$

c) $\mu_+(ER_i) > 0, \;\; \mu_-(ER_i) > 0, \;\; [-\alpha < ER_i < \alpha]$

Let us define the subset of edge responses that belong to category c:

$$ER' = \{ER_i |\; 0 \leq i \leq 7, \;\; \mu_+(ER_i) > 0, \;\; \mu_-(ER_i) > 0\} \tag{6.8}$$

where $ER' \subseteq ER$. ER contains *eight* edge responses, and ER' is the subset of those edge responses in the fuzzy interval $[-\alpha, \alpha]$. For a given pixel, let us define the number of the elements in the subset ER' as $k = |ER'|$, $0 \leq k \leq 8$. Given ER and ER' of a particular pixel, 2^k different 8-bits binary codes can be built as follows:

- If an edge response ER_i belongs to *category a*, ($\mu_+(ER_i) = 1,\ \ \mu_-(ER_i) = 0$), then all the 2^k codes will have '1' in position i of the binary code.
- If an edge response ER_i belongs to *category b*, ($\mu_+(ER_i) = 0,\ \ \mu_-(ER_i) = 1$), then all the 2^k codes will have '0' in position i of the binary code.
- The remaining k edge responses belong to *category c*, and the k bits associated with these k edge responses will be assigned different codes from $\underbrace{000\ldots0}_{\text{k- bits}}$ to $\underbrace{111\ldots1}_{\text{k- bits}}$ in the 2^k binary codes to be built.

Fig. 6.2 presents an example of the calculation of FLDP with the threshold $\alpha = 100$. Both $ER_2 = -10$ and $ER_3 = 30$ are located in the fuzzy interval $[-\alpha, \alpha]$. In this example, $k = 2$, so 2^2 FLDP codes are computed as follows:

- $ER_{0,1,6}$ belong to category b ('0' will be assigned to the associated positions).
- $ER_{4,5,7}$ belong to category a ('1' will be assigned to the associated positions).
- $ER_{2,3}$ belong to category c, so *four* 2-bit combinations: '00', '10', '01' and '11', are assigned to the associated positions of $ER_{2,3}$ in the *four* 8-bits codes.

The critical parameter of FLDP is the selection of the proper value of threshold α. The role of α in the generation of FLDP codes can be explained as follows:

- If α is big, most of the edge responses will belong to category c, and the number of fuzzy cases (i.e. k) for each pixel will be high (on the limit, 256).
- If α is small, most of the edge responses will belong to categories a and b, and the number of fuzzy cases for each pixel will be low (in the limit, 1).
- If $\alpha = 0$, the calculations will be performed in the crisp space (i.e. there is no fuzzy interval). Consequently, the crisp sets of Eq. 6.2 and Eq. 6.3 will be used to calculate the binary codes. If the edge response is *positive*, '1' will be assigned to its associated position in the binary code. If the edge response is *negative*, '0' will be assigned to its associated position in the binary code.

To find the best value of α, a grid search procedure is used. In this chapter, α is allowed to vary in the range $10^2 \leq \alpha \leq 10^3$. For each pixel, a set of m 8-bits binary codes ($1 \leq m \leq 256$) with values between 0 and 255 is generated. Each of these 8-bits codes $C = [c_0 c_1 \ldots c_7]$ may be assigned a particular weight w_c, depending on the degree of the membership function of the associated edge response to the

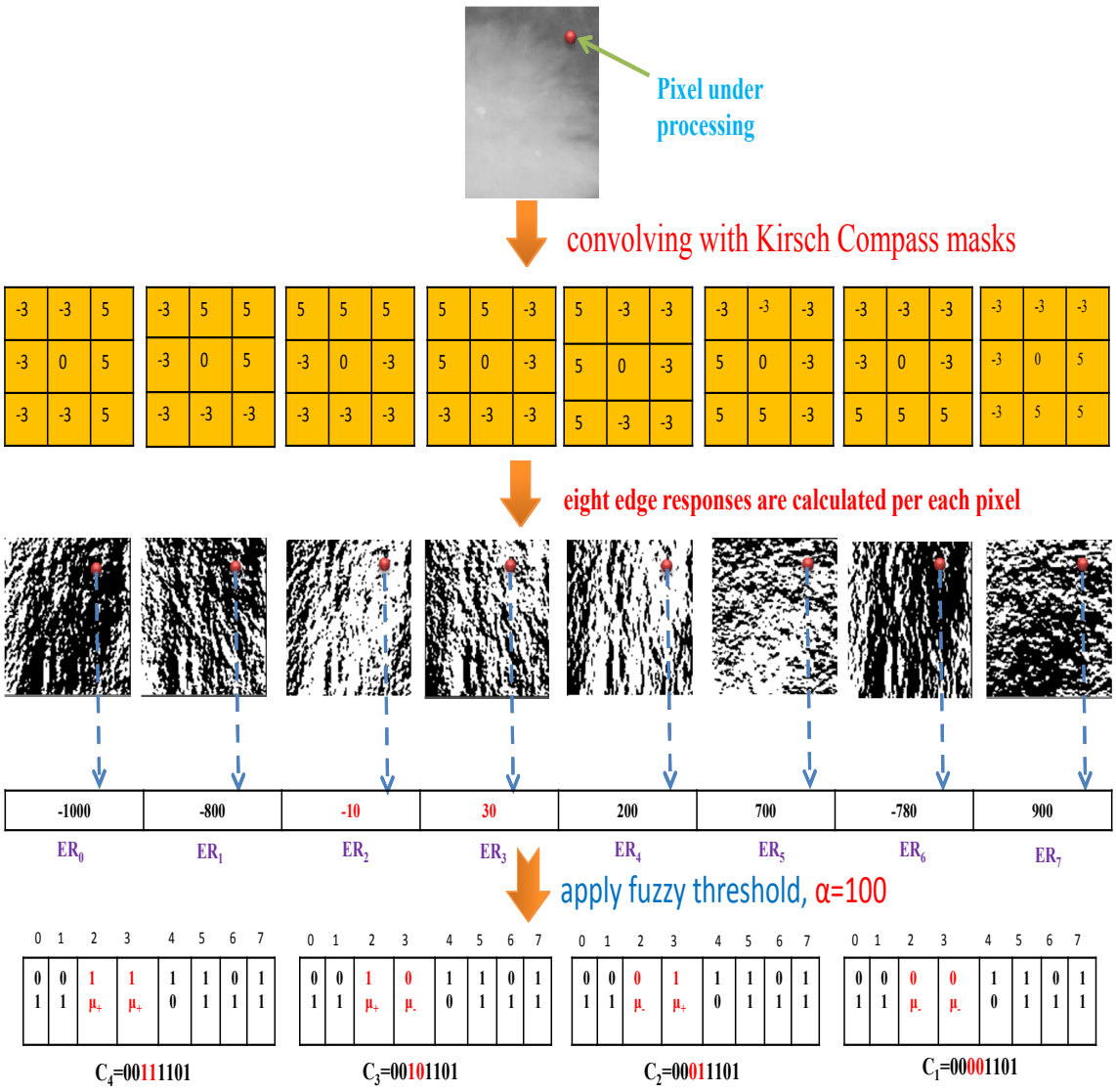


Figure 6.2: Example of the calculation of FLDP codes

positive or negative fuzzy sets (depending on whether the bit in the code is '1' or '0', respectively). Given a binary code C, its weight w_c can be computed as follows:

$$w_c(C) = \prod_{i=0}^{7} c_i.\mu_+(ER_i) + (1 - c_i).\mu_-(ER_i) \tag{6.9}$$

Given a particular pixel with a set of edge responses from which 2^k codes have been generated, it can be proved that the addition of the weights of these 2^k codes is 1. Recalling the example in Fig. 6.2, we can use Eq. 6.6 and Eq. 6.7 to calculate the degree of membership of ER_2 and ER_3 to the positive and negative fuzzy

sets as follows: $\mu_+(ER_2)$=$\mu_+(-10)$=0.45, $\mu_-(ER_2)$=0.55, $\mu_+(ER_3)$=$\mu_+(30)$=0.65, $\mu_-(ER_3)$=0.35. The weight of each fuzzy code can be calculated using Eq. 6.9, $w_c(C_1)$=$0.55 \times 0.35 = 0.1925$, $w_c(C_2)$=$0.35 \times 0.45 = 0.1575$, $w_c(C_3)$=$0.55 \times 0.65 = 0.3575$, $w_c(C_4)$=$0.45 \times 0.65 = 0.2925$. It is clear that the addition of the weights $(0.1925 + 0.1575 + 0.3575 + 0.2925)$ of the fuzzy codes of the pixel given in Fig. 6.2 equals 1.

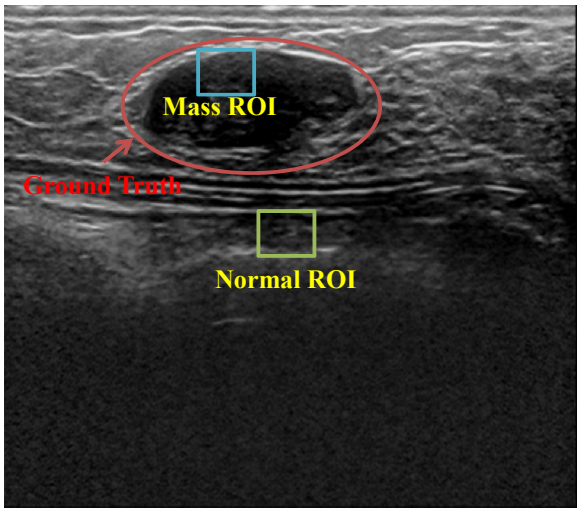
The codes associated with a pixel may be represented graphically in a histogram in which the x-axis is the decimal value of each code (0-255) and the y-axis is the weight associated with that code. Given a grey level ROI, the complete FLDP histogram is computed by adding the weights of all the pixels in the input ROI (Ahonen and Pietikäinen, 2007).

6.3 Experimental results and discussion

In this study, we claim that FLDP is good at characterizing different tissues in breast images. Given a set of normal and mass ROIs, FLDP is extracted from each ROI and fed into an SVM classifier. Then, the trained model is used to classify a query ROI as mass or normal.

A set of 267 breast BUS images is used. The images were collected from 267 patients at the UDIAT Diagnostic Centre of Sabadell (Spain) using a Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz). Of the images, 104 are normal and 163 contain masses. The BUS database contains the GT of the lesions that appear in the abnormal image. We also use the mini-MIAS database (X-ray images) in our experiments (Suckling et al., 1994).

With the BUS database, 32×32 pixel ROIs were extracted (Keramidas et al., 2011). A total of 107 mass ROIs and 300 normal ROIs were extracted. To generate the ROIs we followed the procedure given in (García-Manso et al., 2013). Fig. 6.3 presents an example of the normal and mass ROI generation in breast ultrasound images. With the mini-MIAS database, 109 mass ROIs were extracted from the mass mammograms and 203 normal ROIs were extracted from the normal mammograms. The extracted



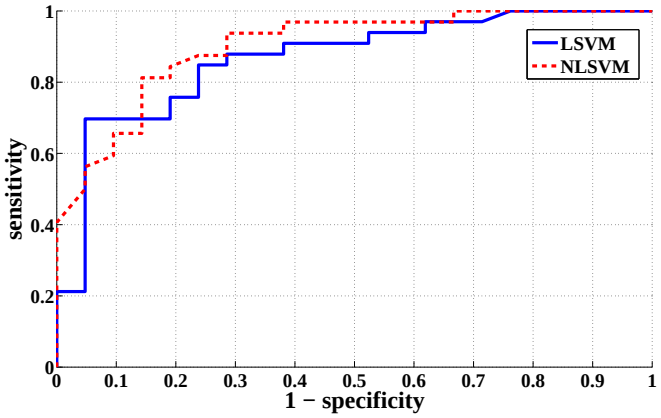
(a)

Figure 6.3: ROI generation of BUS images

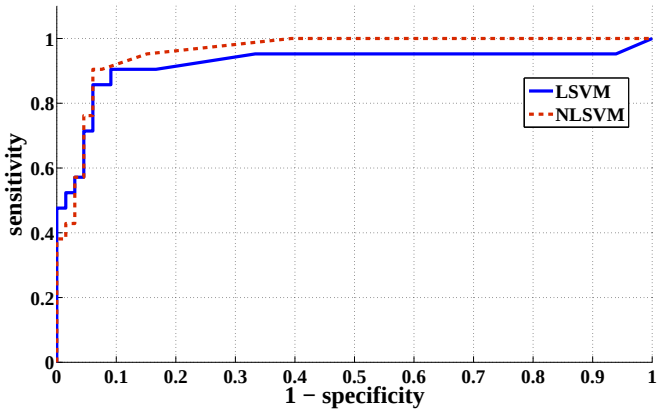
ROIs were of different sizes, so they were resized into a fixed template (in this chapter, 75×75 pixels). Fig. 6.4 (a) shows the ROC curves of BUS ROIs, while Fig. 6.4 (b) shows the ROC curves of X-ray ROIs, with the LSVM and the NLSVM classifiers. With the two datasets, the AUC value is best with the NLSVM classifier.

As mentioned in Section 2, the critical parameter of the proposed descriptor is the threshold α. A *grid search* procedure is used to find the value of α which yields the best AUC value. In our experiments, the optimum value with the X-ray ROIs is 900, whereas with the BUS ROIs it is 750.

To assess the performance of the proposed descriptor, the best mass/normal classification results of the proposed descriptor were compared with some of the-state-of-the-art methods. Table 6.1 presents the results of mass/normal breast tissue classification with the FLDP descriptor applied to the mini-MIAS database, as well as the results of mass/normal breast tissue classification with FLBP (Keramidas et al., 2011), LBP (Oliver et al., 2007), RLBP (Chen et al., 2013), HOG (Pomponiu et al., 2014), LDP (Jabid et al., 2010), MLDP (Mohamed et al., 2014), Gabor (Zheng, 2010) and GLCM features (Soltanian-Zadeh et al., 2004). These descriptors are calculated from the extracted ROIs, and then classified with the same procedure used with FLDP (all descriptors are normalized to unit length). Table 6.1 shows that



(a)



(b)

Figure 6.4: ROC curves of mass/normal breast tissue classification using FLDP with (a) BUS and (b)X-ray datasets

MLDP, Gabor and GLCM features produce the worst AUC values, which indicates that they do not produce a robust description for the breast tissues in the X-ray images. Table 6.1 also shows that Gabor's features produce the worst AUC value with the BUS ROIs.

According to the experiments, FLDP produces the best results with the SVM classifiers. The other descriptors have considerable problems in characterizing the breast tissues particularly in noisy images or dense breasts. For instance, LBP and FLBP assign the same binary code to a pixel in a *tumorous* region and another pixel in a *normal dense* region. This happens when the values of all the neighbours are *higher/smaller* than the value of the centre pixel of a local neighbourhood.

Table 6.1: Comparison between the AUC values of mass/normal breast tissue classification in BUS as well as X-ray images using FLDP, FLBP, LBP, RLBP, HOG, LDP, MLDP, Gabor, and GLCM features with LSVM and NLSVM classifiers

Method	BUS		Mini-MIAS	
	LSVM	NLSVM	LSVM	NLSVM
FLDP	**0.8665**	**0.9140**	**0.9203**	**0.9412**
FLBP	0.8915	0.8981	0.9010	0.9141
LBP	0.9012	0.8775	0.6978	0.8947
RLBP	0.8398	0.8553	0.9103	0.9228
HOG	0.8810	0.8862	0.7664	0.8874
LDP	0.8350	0.9087	0.8195	0.7050
MLDP	0.8408	0.8910	0.5404	0.5338
Gabor filters	0.7560	0.7388	0.6901	0.6412
GLCM	0.8636	0.8882	0.6803	0.6217

In addition, GLCM features depend on the co-occurrence matrix which calculates the number of the pixels with the same intensity at a particular offset (distance and angle). Unfortunately, a similar co-occurrence matrix is produced for a tumorous ROI and a normal ROI in a dense breast region. Moreover, the problem of HOG is the selection of the cell size and the number of cells per block. If the block size is unsuitable, the same HOG descriptor will be produced for a dense normal block and a tumorous block leading to a high number of false detections.

The key advantage of FLDP is the encoding of the edge responses of each pixel using fuzzy logic. Indeed, calculating the edge responses in eight different directions leads to the good characterization of the micro-patterns in a particular ROI. Consequently, if a micro-pattern is missed in one direction, it can be captured in another. In this way, FLDP describes the shapes, margins, spots, edges, corners, junctions and other structures of different tissues in a breast region. Unlike the methods used in the comparison, the use of fuzzy logic provides a range of uncertainty, which makes FLDP able to generate binary codes that compensate the effect of deformations (because of compression), breast density variation and noise. In addition, the histogram that accumulates the weights of the FLDP codes increases the discrimination ability of FLDP, because it encodes both the local information of each pixel as well as the global information of a particular ROI.

6.4 Conclusion

In this chapter, we have proposed FLDP for breast tissue characterization. It properly discriminates between mass and normal tissues in both dense and fatty breasts. FLDP describes each pixel in a given image by its edge responses and makes use of fuzzy membership functions. We have used a breast BUS database and the mini-MIAS database in the experiments. In addition, LSVM and NLSVM classifiers are used to demonstrate the effectiveness of FLDP at discriminating between mass and normal tissues. The results show that the proposed descriptor gives better results than some of the state-of-the-art descriptors.

In the next chapter we propose using super-resolution approaches to improve the performance of texture analysis methods when applied to tumor classification in BUS images.

CHAPTER 7

Breast tumor classification in ultrasound images using texture analysis and super-resolution methods

7.1 Introduction

Although mammography is the most commonly used screening method, breast ultrasound (BUS) has been regarded as a powerful adjunct to mammography for women who have dense breasts, as they permit to identify small cancers that do not appear on mammograms (Scheel et al., 2015). BUS images are also used to discriminate benign tumors from malignant ones.

CAD systems are widely used to analyze masses in BUS images. A CAD system

generally consists of three main components: segmentation of the ROI, feature extraction and classification (Jalalian et al., 2013). Texture features are commonly used to characterize benign and malignant tumors. Unfortunately, BUS images suffer from several artifacts, such as speckle noise, which may distort the appearance of local structures and degrade the performance of texture analysis methods.

In the literature, several CAD systems based on texture analysis methods have been proposed in the last years to discriminate between benign and malignant tumors in BUS images. In (Yang et al., 2013), each BUS image was decomposed into multiple ranklet images. GLCM texture features were extracted from each ranklet image and then sent to a SVM. They used the AUC to evaluate their method. With three BUS datasets, AUC values were 0.918, 0.943 and 0.934. In (Gómez et al., 2012), GLCM texture features were extracted from each ROI and then input into a Fisher discriminant analysis classifier. The best AUC of this method was 0.87. In (Ding et al., 2012), each BUS image was first divided into non-overlapping subregions, and then GLCM features were extracted from each subregion and input into a SVM to define a rough ROI. For each pixel in the ROI, four GLCM features were extracted and a self organizing map was used to construct bags of visual words. An AUC of 0.96 was obtained.

Shi et al. (2010) used GLCM and fractal features with a fuzzy SVM to classify BUS images into benign or malignant, with an AUC of 0.964. In (Lo et al., 2015a,b), BUS images were transformed into the ranklet domain and then GLCM and speckle features were extracted. With a binary logistic regression, an AUC of 0.83 was achieved. In (Flores et al., 2015), the authors selected the best feature set from 26 morphological and 1465 texture features from BUS images. The best classification result was an AUC of 0.942 with a set of five morphological features and the local Fisher discriminant analysis classifier. In (Cai et al., 2015), the authors combined the phase congruency with local binary pattern (PCLBP) to discriminate between benign and malignant tumors. They extracted PCLBP from manually defined ROIs. With SVM, an AUC of 0.894 was achieved. In (Uniyal et al., 2015), the authors extracted texture features from radio-frequency time series and from manually defined ROIs to

generate malignancy maps. They obtained AUC values of 0.86 with SVM and 0.81 with RF.

As shown above, previous work has focused on methods extracting texture features from a single BUS image for each breast, which mainly suffers from speckle noise and low contrast. These factors may degrade the discrimination power of the applied texture methods. To address this problem, this chapter proposes the use of an image super-resolution (SR) approach prior to texture analysis. This super-resolution approach extracts a high-resolution (HR) image from a set of low-resolution (LR) images. The HR image obtained provides not only a better visual appearance but also more details with less noise and artifacts; thus, texture analysis methods can accurately characterize breast tumors. We have used five texture methods to demonstrate the effectiveness of our approach. These methods have been included in our CAD system, which consists of four successive stages. In the first stage, an HR BUS image is obtained from a set of BUS images of the same tumor. In the second stage, the ROI is segmented from the HR BUS image. In the third stage, texture features are extracted. Finally, using these features, the segmented ROI is classified into benign or malignant.

7.2 Methods

Fig. 7.1 presents the proposed CAD system. It has two phases: training and testing. In the training phase, the SR algorithm reconstructs an HR image from a set of LR images extracted from an input BUS image sequence. Then, the system extracts texture features from the ROIs that are segmented from the HR images. Finally, the extracted features are fed into a classifier to build a model. The model is trained to discriminate between benign and malignant ROIs. Note that in the training phase the labels (benign or malignant) of the input image sequences are known. The testing phase is similar to the training phase. The only difference is that we feed the extracted features into the model obtained in the training phase. The model predicts the label of the input image sequence (label ‘1’ for malignant cases and label ‘0’ for

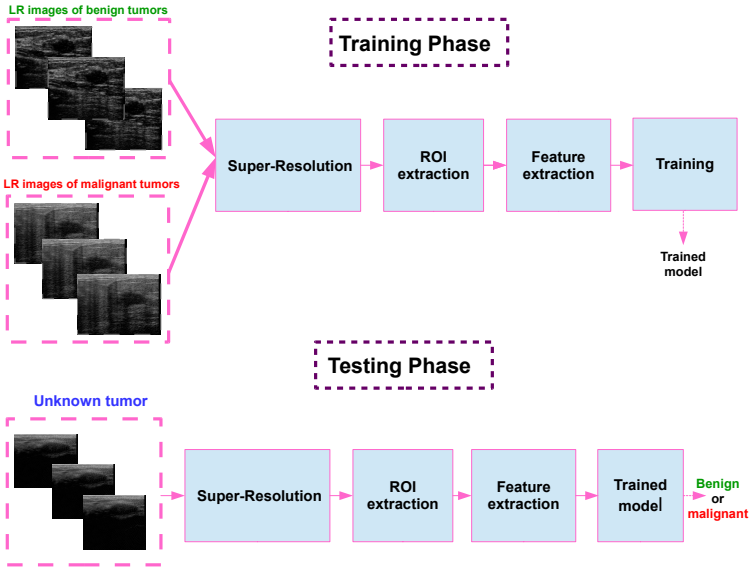


Figure 7.1: The training and testing phases of the proposed CAD system

benign cases). In the sections below, we provide more details.

7.2.1 Input images

A database of 31 malignant and 28 benign BUS image sequences is used to evaluate the proposed CAD system. This dataset is part of a clinical database of ultrasonic radio frequency strain imaging data that was created by the Engineering Department of Cambridge University. The complete database is available at http://mi.eng.cam.ac.uk/research/projects/elasprj/.

7.2.2 Computing the HR BUS image

Given a set of LR images of the same scene, which are degraded versions of an HR image X, the following degradation process models the observation of LR images:

$$Y^k = DHF^kX + V^k \tag{7.1}$$

In this expression, Y^k is the k^{th} observed LR image, X is the HR image, D is a down-sampling operator, H is a blurring operator, F is the motion operator and V^k is the noise in the k^{th} frame, $k = 1, \ldots, N$, where N is the number of LR images.

As we can see in Eq. 7.1, the SR problem is an inverse problem. The HR image $\hat{X}$ can be estimated from a set of LR images by minimizing the following cost function:

$$\hat{X} = arg \min_{X} \left[\sum_{k=1}^{N} \rho(Y^k, DHF^k X) + \lambda \Gamma(X) \right] \tag{7.2}$$

In Eq. 7.2 ρ is a similarity cost function, Γ is a regularization function and λ is a regularization parameter.

For the similarity cost function, $L1$- and $L2$- norms are widely used to fuse LR images (Farsiu et al., 2004; Elad and Hel-Or, 2001). The performance of the $L1$- and $L2$-norms was compared in (Farsiu et al., 2004), and the $L1$-norm proved to be more robust. In this study, then, we use the $L1$-norm as a similarity cost function. It can be calculated as follows:

$$\rho(Y^k, DHF^k X) = ||Y^k - DHF^k X||_1 \tag{7.3}$$

The bilateral total variation (BTV) is used for regularization. It can be defined as

$$\Gamma(X) = \sum_{l=-p}^{p} \sum_{m=-p}^{p} \alpha^{|l|+|m|} ||X - S_x^l S_y^m X||_1 \tag{7.4}$$

where α is a scalar weight ($0 < \alpha < 1$) used to apply a spatially decaying effect to the summation of the regularization term. S_x^l is a shifting operator in the horizontal direction by l pixels, whereas S_y^m is a shifting operator in the vertical direction by m pixels.

Parameter settings. We set the parameters of the L1-norm SR method (Farsiu et al., 2004) as follows. For the number of LR images N, we assessed several values (3, 5, 7 and 9). The LR images were selected from the middle of each input sequence. The resolution increment factor was set to 2. The motion parameters were estimated blindly from the LR images using Lucas-Kanade affine motion model (Bouguet, 2001). The spatial window size in the gradient of the bilateral-filter was set to 5×5. The size of the point spread function (PSF) was set to 5×5. This value

was common to all frames. The exponential decay coefficient (α) and the steepest descent coefficient were set to 0.2 and 0.5, respectively. The maximum number of iterations was set to 10. However, when the cost function converged, the algorithm automatically stopped.

7.2.3 ROI extraction

We used the algorithm proposed in (Shan et al., 2008) to automatically extract the ROIs that contain the tumors. This method automatically selects the seed point for BUS image segmentation. It consists of four steps: speckle reduction, iterative threshold selection, removing of the boundary-connected region and ranking of the regions. After selecting the seed point, a region growing method is used to obtain a preliminary lesion boundary. Based on the region growing result, the minimum bounding rectangle containing the segmented mass is extracted, which constitutes the desired ROI.

The ROI extraction method used requires a few parameters to be set. We used the same values as in (Shan et al., 2008). The number of iterations of speckle reduction was 10. The ratio of the number of foreground points in the iterative threshold step was 0.1. The region growing method used 8-neighborhood connectivity. To understand the role of SR and ROI extraction stages in the proposed CAD system, Fig. 7.2 shows benign and malignant ROIs before and after using the SR algorithm. As we can see, the boundaries of the tumor are clearer with SR. Thus, the texture features used in the feature extraction stage can precisely characterize those boundaries, which is important if the benign and malignant cases are to be discriminated (Mendelson et al., 2013).

7.2.4 Texture features

In this subsection we briefly explain the five texture analysis methods used in this study: GLCM, LBP, phase concurrency based local binary pattern (PCLBP), HOG and pattern lacunarity spectrum (PLS).

Gray level co-occurrence matrix. In the GLCM the distribution of co-occurring

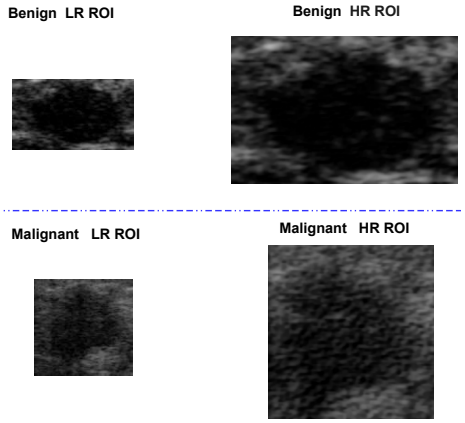


Figure 7.2: ROIs of benign and malignant cases segmented from LR- and HR-BUS images

grey level values in a given direction and at a given distance is computed (Haralick et al., 1973). In other words, the GLCM computes the joint frequencies $p(i, j)$ of pairwise combinations of gray levels i and j separated by distance d along direction θ. If N_g is the number of distinct gray levels in the quantized image, then the size of the GLCM is N_g^2.

Different texture descriptors can be computed from GLCM. To compute GLCM features, we used four orientations (0^o, 45^o, 90^o and 135^o), a distance of 5 and a quantification level of 32. From each GLCM, we calculated 22 texture features and concatenated all of them into one feature vector (its length was $4 \times 22 = 88$ dimensions). The GLCM features and their mathematical expressions are listed in Tables 2.1 and 2.2 of Section 2.3.2.3.

Local binary pattern. In this study we used 3×3 local neighborhoods to generate an LBP descriptor of each ROI (each pixel is represented by eight bits). We computed the histogram with the frequency of uniform LBPs for each ROI. The dimension of the LBP feature vector was 59.

Phase congruency based local binary pattern. Cai et al. (2015) combined the phase congruency (PC) with the local binary pattern (PCLBP) to discriminate between benign and malignant tumors. They extracted PCLBP from manually defined ROIs. PC acts as a line or edge descriptor of the input ROI. Unlike other edge detectors, it is insensitive to image variations due to illumination changes,

blurring and magnification (Štruc and Pavešić, 2009). It has this robustness because it is calculated using multi-scale and multi-orientation responses. Moreover, only the phase information, which is less sensitive to the above variations, is used to calculate it. Assume that R_{st} is a 2D log-Gabor filter at scale s and orientation t, where $s = 1, 2, \ldots, N_s$ and $t = 1, 2, \ldots, N_t$. N_s is the number of scales and N_t is the number of orientations. The convolution of the input ROI with each filter response R_{st} produces two components for each pixel: A_{st}, which is the real part, and $\Im_{st}$, which is the imaginary part. The 2D-PC can be calculated as follows.

$$PC = \frac{\sum_t \sum_s W_t(x)\psi(A_{st}(x)\Delta\Phi_{st}(x) - T_t)}{\sum_t \sum_s A_{st}(x) + \epsilon} \tag{7.5}$$

In this equation, $\Delta\Phi_{st}(x)$ is the phase deviation at pixel x, T_t is the estimated noise energy at orientation t, $W_t(x)$ is a weighting function that weights for the frequency spread, ϵ is used to avoid division by zero and $\psi()$ is a function that rectifies the input as follows: $\psi(X - T)$ outputs $(X - T)$ if $X > T$; otherwise, it outputs '0'. After estimating the 2D-PC form the input ROI, the LBP codes are extracted from the PC image. The resulting features are called PCLBP.

Histogram of oriented gradients (HOG). In this study we used a 3×3 cell size, 4×4 cells for the block size, and a 9-bit histogram. The dimension of the HOG feature vector was 144 ($9 \times 4 \times 4$).

Pattern lacunarity spectrum. Several fractal-dimension based methods have been proposed to analyze textures of medical images (Chen et al., 2005; Cabral and Rangayyan, 2012). Their basic idea is to use multiple fractal dimensions to summarize the spatial distribution of image patterns. Lacunarity is a specialized term in fractal geometry referring to a measure of how patterns fill the space. It can characterize spatial features and describe multi-fractal and even non-fractal patterns. In (Quan et al., 2014), pattern lacunarity spectrum (PLS) was proposed for texture classification. In this study we evaluate the performance of PLS in classifying benign and malignant tumors in BUS images.

To calculate the PLS descriptor, the following four sequential steps are performed:

1. Calculate LBP label images $(Z_1, Z_2, \ldots, Z_N)$ using different (P_i,R_i) pairs,

where P_i denotes the shape of the neighborhood, and R_i defines the size of the neighborhood, $i = 1, 2, \ldots, N$.

2. Determine binary images $B_{i,j}(x, y)$ from each LBP label image Z_i with $j = 1, 2, \ldots, P_i + 2$ as follows.

$$B_{i,j}(x, y) = \begin{cases} 1 & if\ Z_i(x, y) = j \\ 0 & otherwise \end{cases} \tag{7.6}$$

3. Calculate lacunarity-related features $[D(B_{i,j}), L(B_{i,j})]$ from each binary image $B_{i,j}(x, y)$.
4. Concatenate all lacunarity-related features into one vector (called PLS).

To compute the lacunarity-related features $D(B_{i,j})$ and $L(B_{i,j})$, the lacunarity is determined first. The simplest method for calculating lacunarity on a binary image $B_{i,j}$ is the sliding box method (Allain and Cloitre, 1991), where an $s \times s$ box is slid through the binary image. The number of mass points within the box at each position is calculated. Then, a histogram $X_s(n)$ is built; n denotes the number of mass points falling into the box, and $X_s(n)$ is the number of boxes containing n mass points. The lacunarity at scale s can be calculated as follows:

$$A_s(B_{i,j}) = \frac{E[(X_s)^2]}{(E[X_s])^2} \tag{7.7}$$

In (Mandelbrot, 1983), it is stated that lacunarity exhibits power law behaviors with respect to its scale s as follows:

$$A_s(B_{i,j}) \propto \left(\frac{1}{s}\right)^{D(B_{i,j})} \tag{7.8}$$

By taking the logarithm on both sides of the above equation, the following equation is obtained

$$\ln A_s(B_{i,j}) = D(B_{i,j}) \ln s + L(B_{i,j}) \tag{7.9}$$

Finally, the linear least squares fitting technique is used to estimate $D(B_{i,j})$ and $L(B_{i,j})$. In this study we used the same parameter values as in (Quan et al., 2014).

The scale range for estimating lacunarity was [2,14] with a step of 1. The parameters (P_i, R_i) were $(4,1),(16,2),(16,3),(8,5)$ and $(16,5)$.

7.2.5 Classification stage

In this study, we used an RF with 30 trees, which was the optimal number of trees that stabilized the out-of-bag error. We used the leave-one-out cross validation (LOOCV) technique to generate the training and testing sets (LOOCV is recommended for small size datasets). The ROC curve was obtained for each experiment, and the area under the ROC curve (AUC) was used to measure the performance of each texture analysis method.

7.3 Results

Table 7.1 shows the AUC values obtained for each texture analysis method using HR images computed from different sets of LR images. With all texture methods, AUC values are best when the number of LR images is set to 5. With 5 LR images, the AUC value is best with the HOG descriptor (0.989) followed by LBP (0.950). With nine LR images, LBP gives the worst AUC value (0.541). This demonstrates that the number of LR images affects the quality of the reconstructed HR image, so it also affects the performance of the proposed CAD system.

Physicians use the shape of the boundary of the lesion to discriminate between benign and malignant lesions. The boundaries of benign lesions usually have a regular shape, while malignant lesions are usually irregular. Thus, an accurate characterization of the boundaries of the lesions leads to good classification results. HOG characterizes the edges in each ROI and works well in the case of illumination changes. Given a ROI extracted from an HR image, HOG produces an accurate characterization for the edges. This is the main reason why the AUC values of HOG are larger than with the other methods.

To understand why the HOG descriptor leads to the best classification results with SR, we can visualize its output. Fig. 7.3 shows the HOG features extracted from

Table 7.1: AUC values of the proposed CAD system when using different texture methods and different numbers of LR images to compute HR images

Methods/N_{LR}	**3**	**5**	**7**	**9**
LBP	0.667	0.950	0.595	0.541
HOG	0.786	**0.989**	0.777	0.802
PCLBP	0.660	0.923	0.638	0.652
GLCM	0.711	0.888	0.602	0.725
PLS	0.510	0.857	0.560	0.706

a ROI before and after the SR step. To make a clear visualization, we construct a HOG descriptor with a cell size of 8×8 pixels. In the case of HR-ROI, HOG produces a good description for the whole boundary of the tumor, whereas in the case of LR-ROI, it produces a poor description and the resulting output disregards a large amount of information about the boundary. The accurate HOG description achieved with SR helps the classifier to discriminate between benign and malignant cases, leading to excellent classification results.

As explained in the introduction, the authors of the studies of a similar nature to

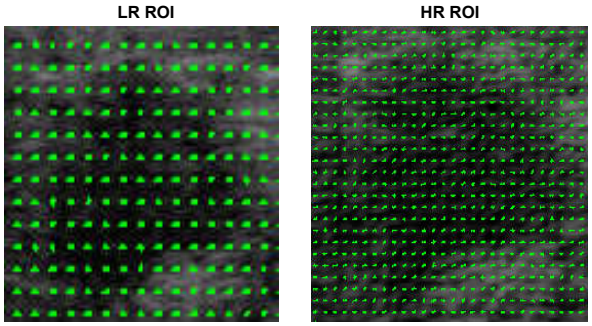


Figure 7.3: Visualizing HOG features in LR-ROI and HR-ROI

this one used a single image for each breast while the proposed CAD system uses an HR image computed from multiple LR images. To compare the results of the proposed system with those of the other studies, the second column in Table 7.2 shows the performance of the texture methods when a single image is used from each input image sequence (with no pre-processing (NP)). In this experiment, we pick one LR frame from the middle of each input sequence (the middle of the LR sets that we used to construct HR images). Then, we extract texture features from the selected frame and input them into a RF classifier to discriminate between

benign and malignant cases. We call this approach the *one image system* (OIS). HOG gives higher AUC values (0.828) than the other texture methods. GLCM gives the lowest one (0.464). As the GLCM counts the co-occurrence of the pixels, any small noise will lead to an inaccurate characterization. Therefore, the texture features calculated from each GLCM do not properly describe the lesions of BUS images. In addition, LBP compares the intensity values of the central pixels in a small neighborhood with the intensities of its neighbors. Thus, noise may push LBP to calculate the same binary codes for two different neighborhoods leading to an inaccurate description of the ROI. As a result, noise also affects the performance of LBP leading to a low AUC of 0.563.

We also analyze the effect of pre-processing operations when using OIS and compare its results with the proposed system. To do so, we adopted the median filter (MF) and histogram equalization (HE) techniques. In the case of MF, we replaced each pixel by the median of a 3×3 neighborhood centered around it. As shown in Table 7.2, both MF and HE fail to improve the performance of the texture methods. The best AUC of OIS is achieved with HOG with no pre-processing.

Table 7.2: AUC values of OIS with NP, MF and HE operations

Methods	NP	MF	HE
LBP	0.563	0.518	0.492
HOG	**0.828**	**0.787**	**0.807**
PCLBP	0.578	0.559	0.572
GLCM	0.464	0.402	0.412
PLS	0.619	0.536	0.540

To illustrate the discrimination capability of the proposed system, we computed the malignancy score of each case with the new SR-based approach and the OIS system using HOG features. The scores of the RF were used to compute the malignancy score. Note that RF scores indicate the likelihood that a label comes from the malignant class. We ranked the cases according to their malignancy score in descending order. A perfect classification system should assign a malignant case with a rank from 1 to 31, and a benign case with a rank from 32 to 59. Fig. 7.4 shows some malignant and benign cases. The proposed system assigned the *first*

rank to the malignant case shown in Fig. 7.4(a) while it assigned *rank 25* to the malignant case shown in Fig. 7.4(b). In turn, the OIS system assigned *rank 39* to the case shown in Fig. 7.4(a) and *rank 44* to the case shown in Fig. 7.4(b). Thus, OIS wrongly classified both cases as benign tumors. The OIS system assigned *ranks 26* and *27* to the benign cases shown in Fig. 7.4(c) and 7.4(d), respectively. These ranks make the OIS system wrongly classify these cases as malignant tumors. In contrast, the SR-based system assigned the last rank (*59*) to the benign case shown in Fig. 7.4(d) and *rank 37* to the case shown in Fig. 7.4(c). As a result, the proposed system correctly classified these cases as benign tumors. The previous cases were better discriminated using SR, which demonstrates the effectiveness of the proposed method.

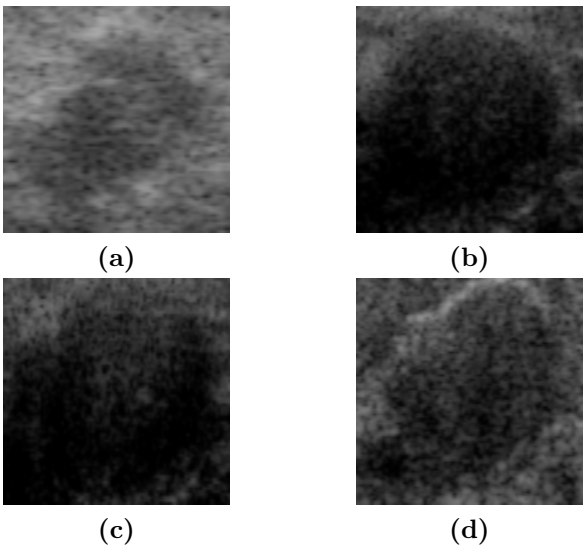


Figure 7.4: Comparing the performance of the SR-based approach and the OIS system. (a-b) malignant, (c-d) benign cases

Statistical analysis. The AUC values in Table 7.1 are better than those in Table 7.2, which indicates that the proposed CAD system improves the classification results of all texture methods. It is interesting to determine the statistical significance of the differences of performance between the proposed CAD system and the OIS in terms of the AUC. To do so, we analyzed the statistical significance of the difference in AUC values of each texture method considering the proposed CAD system and

using 5 LR images and the OIS with NP. Then we used *Welch's* t-test (significance level < 0.05) to determine the difference in AUC values. The normality of the distributions of the AUC values was assessed by means of bootstrapping and the Shapiro-Wilk test. We found that all the AUC values follow a normal distribution. We also used the *F-test* to check if two different groups had the same variance: the variances between all pairwise groups were unequal. Note that, in our analysis, the number of comparisons equals 5 and the significance level α of a single experiment equals 0.05; so, according to the Bonferroni correction (Curtin and Schulz, 1998), the actual significance level is 0.01 (0.05/5). Table 7.3 shows the statistical analysis of the AUC values obtained. In this table, a p-value lower than 0.01 indicates statistical significance. As can be seen in Table 7.3, the AUC values of the proposed CAD system with all texture methods are significantly better than the values of the OIS.

Table 7.3: Comparison between the proposed CAD system and OIS in terms of the AUC (Statistically significant differences are shown in bold)

Proposed vs. OIS	**Best AUC of the proposed CAD**	**Best AUC of the OIS**	**p-value**
LBP	0.950	0.563	**<0.0001**
HOG	0.989	0.828	**<0.0001**
PCLBP	0.923	0.578	**<0.0001**
GLCM	0.888	0.464	**<0.0001**
PLS	0.857	0.619	**<0.0001**

7.4 Discussion

BUS images suffer from speckle noise, shadowing and other artifacts. BUS artifacts are the set of structures in ultrasound images that do not have corresponding anatomical structures. Artifacts usually appear when a BUS image is displayed; they are produced by the physical properties of ultrasound themselves. Artifacts can be classified into four main categories: missing structures, degraded images, falsely perceived objects and structures with a mis-registered location (Sehmbi and Perlas, 2015). Obviously, artifacts may yield unwarranted clinical intervention.

These factors also degrade the performance of BUS-CAD systems that use texture analysis methods in their feature extraction stage. Some texture analysis methods, such as GLCM and LBP, are quite sensitive to the aforementioned factors. Thus,

enhancing BUS images before feature extraction can improve the overall performance of CAD systems. In this study, we used SR to enhance the input images of BUS-CAD systems. As we have shown, texture features calculated from SR images can accurately characterize breast lesions.

In general, CAD systems include a pre-processing stage to prepare BUS images. For instance, Veeramani and Muthusamy (2015) use an adaptive median filter to minimize the effect of noise and preserve useful details. Lo et al. (2015a,b) use the ranklet transformation to reduce the effect of image variations. Cai et al. (2015) compared the effect of three pre-processing operations on PCLBP: contrast improvement, gamma correction and histogram equalization. They showed that PCLBP produced comparable results with the three pre-processing operations. In turn, we used SR in this study. As shown in Fig. 7.4, instances of malignant and benign cases were better discriminated using SR. The proposed system obtained high malignancy ranks for malignant cases and low malignancy ranks for benign ones. In contrast, the OIS system obtained high malignancy ranks for several benign cases and low malignancy ranks for several malignant ones. In addition, statistical analysis showed that the AUC values of the proposed CAD system were significantly better than those of OIS. These findings show the discrimination power of the proposed system and that the use of SR improves the performance of state-of-the-art CAD systems.

In this study, we used SR in the pre-processing stage and studied its effect on five texture analysis methods: LBP, HOG, PCLBP, GLCM and PLS. The AUC value was the highest with HOG. Indeed, several studies have used the HOG descriptor with ultrasound images. Kawahara et al. (2014) proposed an approach to predict the probability that an input frame contains tumorous tissue using RF. The accuracy of the approach was 85.5%. In (Veeramani and Muthusamy, 2015) the HOG descriptor is used to detect abnormalities in ultrasound lung images. In (Agarwal et al., 2013), HOG is used to discriminate parasternal long axis (PLAX) and short axis (SAX) B-mode echocardiograms with an accuracy of 98%. Unlike the aforementioned methods, the proposed CAD system uses the SR in its pre-processing stage and

classifies between benign and malignant tumors, achieving an AUC of 0.989 with HOG.

In this chapter, we focused on improving the performance of the texture methods when classifying malignant and benign lesions. Alternatively, other works like (Zhou et al., 2015) focused on characterizing the shape of the tumor contour using a set of shape features. Zhou et al. (2015) have considered the fact that the top half part of the tumor contour is less affected by posterior acoustic shadowing (PAS). Half-contour features were used to classify benign and malignant breast tumors, with a final AUC of 0.81. Obviously, the performance of (Zhou et al., 2015) can be improved if it is combined with the proposed system. Specifically, SR can be used to reconstruct a HR image from a set of LR images for the same lesion. Given the HR image, the calculated shape features will be more descriptive and give more accurate classification results.

7.5 Conclusion

This chapter proposes the use of image super-resolution to improve the performance of texture analysis methods when applied to benign/malignant tumor classification in breast ultrasound images. Given an ultrasound image sequence, we select a set of LR frames and then use them to reconstruct a HR image. The texture features extracted from the HR images properly describe lesions. In this way, the performance of CAD systems that use texture analysis methods in the feature extraction step can be improved. The proposed CAD system achieved an AUC of 0.99 with 5 LR images. Statistical analysis showed that the AUC values of the proposed CAD system were significantly higher than the values of OIS schemes. In general, SR as a pre-processing step need not be restricted to CAD systems that use texture analysis methods; it can also be used with other feature extraction approaches, such as shape descriptors. In the next section, we propose an automatic method for detecting nipples in thermograms.

Part IV

Analysis of breast cancer in infrared images

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

CHAPTER 8

Automatic nipple detection in thermograms using image processing and anatomical information

8.1 Introduction

Current breast cancer studies show that detecting breast cancer in its early stage may reduce mortality (Siegel et al., 2015). The study of mammographies is the most common method for detecting breast cancer. Besides, the study of breast thermographies has been considered as an adjunct breast cancer screening procedure (Ng and Sudharsan, 2004). Thermograms are infrared images of the human body that can be used to detect breast cancer in its early stage. They can also reveal tumors

in dense tissues, outperforming other modalities such as mammographies. However, the accuracy of thermographies is dependent on such factors as the symmetry of breast temperature and temperature stability.

Infrared imaging can be used for breast cancer screening because the metabolic activity and vascular circulation in pre-cancerous tissue and in the area surrounding a developing breast cancer are always higher than in normal breast tissue. Mammographies, sonographies and magnetic resonance images try to find the physical tumor. In turn, thermograms detect the heat produced by increased blood vessel circulation and metabolic changes associated with a tumor's genesis and growth.

A number of CAD systems have been proposed for analyzing thermograms. A basic step in the use of thermal images is the design of CAD systems because they help technicians to execute medical exams using specified routines and protocols (similar to mammography exams). CAD systems give physicians good support in the analysis and interpretation of the outcomes of thermograms.

In the literature, several studies have focused on analyzing thermograms. Borchartt et al. (2013) presented a review of the literature on image processing and techniques related to the analysis of thermographies for the detection and diagnosis of breast diseases. In (Saniei et al., 2015), a computer method for quantifying the bilateral differences between left and right breasts was proposed. Ali et al. (2015) proposed a simple segmentation approach for extracting the ROI from breast thermograms. They extracted statistical and texture features from the segmented ROI and used a support vector machine classifier to discriminate between normal and abnormal breasts. Krawczyk et al. (2015) analyzed breast thermograms by extracting features describing bilateral symmetries between the two breasts, and then classifying them into benign or malignant. Schaefer et al. (2009) manually segmented the breast region from each thermogram and then extracted a set of statistical features from them. Finally, they used a fuzzy rule-based algorithm to discriminate between benign and malignant cases.

The nipples are an important anatomical landmark in thermograms. The location

of the nipples is invaluable in tasks such as image registration and modality fusion. Indeed, physicians use the location of nipples to find the corresponding masses in breast images that were acquired using different modalities (e.g., a thermogram and a mammogram). For instance, Saniei et al. (2015) manually picked the nipples to register the images of breasts. When analyzing a large number of thermograms, manual nipple detection may be time consuming; in addition, the accuracy of nipple detection varies with user precision. Therefore, the design of an automatic nipple detection method may be very useful in the analysis of thermograms.

Several methods have been proposed for detecting nipples in mammograms (Jas et al., 2013; Chakraborty et al., 2015) or in ultrasound images (Moghaddam et al., 2014; Wang et al., 2014a). However, few solutions have been proposed for detecting nipples in thermograms. Koay et al. (2004) proposed a method for classifying thermograms using an artificial neural network. They briefly mentioned an abridged rule-based method for detecting nipples. However, they did not explain any parameter settings or present an evaluation of the method because the main aim of their work was the classification of thermograms into normal, fibrocystic or cancerous. Unlike this study, in this chapter we propose an unsupervised, automatic, accurate, simple and fast method for detecting nipples in thermograms using image processing operations and anatomical information. Moreover, we propose a novel selection algorithm that selects the nipples from a set of candidates.

8.2 Proposed method

The stages of the proposed method are summarized in Fig. 8.1. As shown, there are three sequential stages: human body segmentation, determination of nipple candidates using adaptive thresholding and detection of the nipples using a selection algorithm.

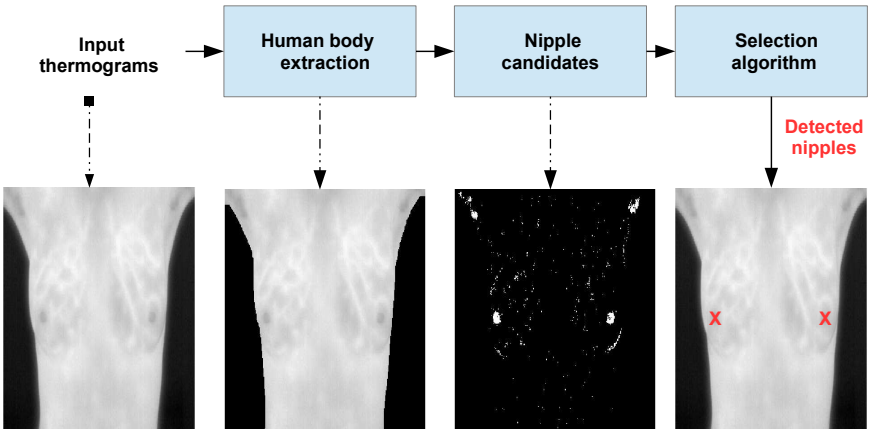


Figure 8.1: Proposed system

8.2.1 Human body segmentation

Thresholding is used to segment the human body from a given image by setting all pixels whose intensity values are above a threshold to a foreground value (one) and all the remaining pixels to a background value (zero). A threshold of 50 is used in this study to generate the human body mask (binary image), which is then smoothed by morphological operations. We applied morphological closing to the binary mask using a disk-shaped structuring element with a radius of 3 to fill the gaps. We then used a morphological dilation with a disk-shaped structuring element with a radius of 10. Fig. 8.2(a) shows an example of an input thermogram and how the generated human body binary mask (Fig. 8.2(b)) is used to segment the human body image (Fig. 8.2(c)).

(a) **(b)** **(c)**

Figure 8.2: Human body segmentation. (a) the input thermogram image, (b) the human body mask, and (c) the segmented image

8.2.2 Determination of nipple candidates using an adaptive threshold

To determine nipple candidates, we process the image that contains the human body by means of *an adaptive thresholding* algorithm. Adaptive thresholding typically takes an image as input and produces a segmented image and the labels of the segmented regions. It is a method that can separate the foreground from the background with non-uniform illumination (He and Yung, 2004).

For each pixel in the image, a threshold has to be calculated. If the pixel value is below the threshold it is set to the background value; otherwise, it is set to the foreground value. To find the local threshold, the algorithm statistically examines the intensity values of the local neighborhood of each pixel. Functions such as mean or median of the local intensity distribution can be used. In Algorithm 1, we present the steps of the adaptive thresholding.

Algorithm 1 Adaptive thresholding algorithm

1: Input: a gray scale image
2: Output: a binary image
3: **procedure** ADAPTIVE THRESHOLDING
4: Convolution of the image with a suitable statistical operator (the median or mean).
5: Subtraction of the original image from the convolved one.
6: Thresholding the difference image with a constant C.
7: Invert the thresholded image.
8: **end procedure**

In this study we used the median operator with a local neighborhood of 15 pixels and C was 0.03. Fig. 8.3 shows nipple candidates extracted from the human body image using the adaptive thresholding algorithm. As shown, most candidates are a long way from the real position of the nipples. Moreover, some regions correspond to external objects, like necklaces, bracelets or wristwatches (e.g., in Fig. 8.3 we can see a necklace on the upper part of the image). Thus, we need a selection algorithm to pick the correct location of nipples. In the subsection below we explain the proposed nipple selection algorithm.

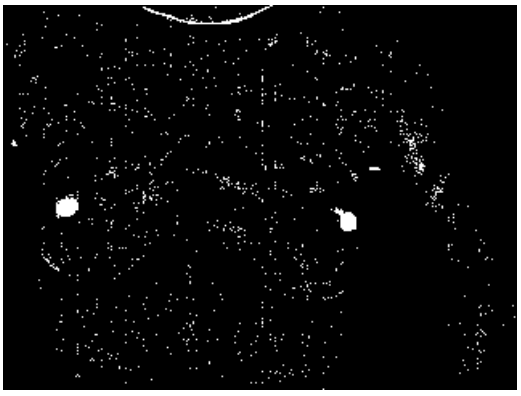


Figure 8.3: Determining nipple candidates using adaptive thresholding

8.2.3 Nipple detection

To select the correct position of the nipples from a set of candidates we applied several selection rules. We used our knowledge of the features of thermograms to design these rules. Breast nipples have some unique anatomical and visual features that can be used to design the selection rules. We used the following facts:

- Nipples lie inside the outer boundary of the human body.
- The human body has only two nipples, which occupy small regions.
- Nipples do not lie in the lowest or uppermost parts of thermograms.
- A thermogram has only two nipples, one on the left and another on the right.
- A breast nipple is approximately circular.

Exploiting the first fact. The human body boundary is extracted by defining the outermost edges in the human body mask. To do so, the rows and columns are scanned twice. Each row of the mask is scanned from the left to determine the right most edge points, and then they are scanned from the right to determine the left most edge points. Each column of the mask is scanned from the top to determine the bottom most edge points and then scanned from the bottom to determine the top most edge points.

Exploiting the second fact. We count the number of pixels in the region of each candidate. If the number of pixels in a given region is less than a predefined threshold N_p, we delete it from the nipple candidates list. In this study N_p was set to 20.

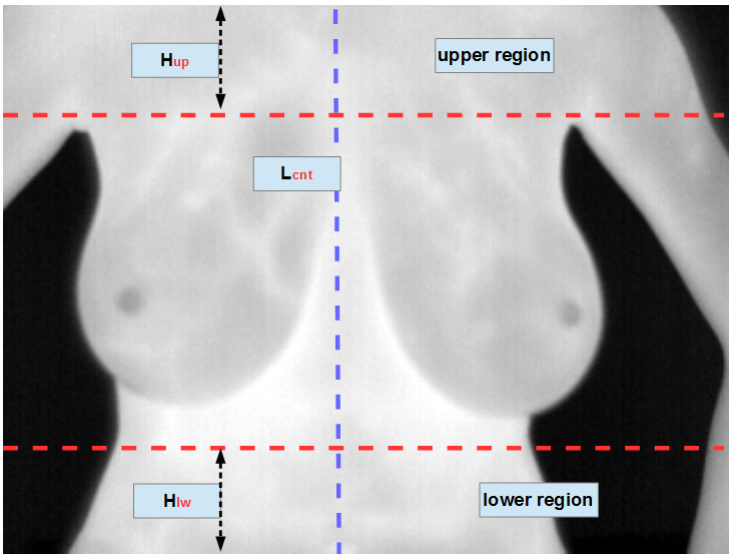
Exploiting the third fact. We generate a binary mask ($mask_{UL}$) to exclude the

regions that lie in the uppermost and lowermost parts of the image. Fig. 8.4 shows the upper region ($region_{up}$) and lower region ($region_{lw}$). Candidate regions that lie in or touch the $region_{up}$ or $region_{lw}$ are excluded from the nipple candidates list. The regions that correspond to $region_{up}$ and $region_{lw}$ in $mask_{UL}$ are set to zero; the rest of the image is set to one. Assume that H_{im}, H_{up} and H_{lw} are the height of the input thermogram, $region_{up}$ and $region_{lw}$, respectively. H_{up} and H_{lw} have been defined as follows.

- H_{up} was set to $0.35 * H_{im}$.
- H_{lw} was set to $0.3 * H_{im}$.

A similar setting was used in (Ali et al., 2015) to segment the breast region.

Exploiting the fourth fact. To determine the left and right regions of the input



(a)

Figure 8.4: Determining the upper region ($region_{up}$), the lower region ($region_{lw}$) and the center-line (L_{cnt})

thermogram, we determine the center-line (L_{cnt}) as shown in Fig. 8.4. L_{cnt} divides the input image into two equal regions: one region includes the left breast and the other includes the right breast.

Exploiting the fifth fact. As nipples are approximately circular, we also use this feature to select the correct nipples from the candidates list (N). In this study we use the *roundness* measure to identify the candidates that are round in shape. To measure the roundness (R) of a given region we used the following metric (Crocker

et al., 1983):

$$R = 4\pi.A/P^2 \tag{8.1}$$

In this equation A and P are the area and the perimeter of the region of a given candidate, respectively. As the value of R approaches 0, it indicates an oblong (non-circular) object. A value of 1.0 indicates a perfect circle. The area of a region is measured simply by counting the number of pixels. Given the locations of the points of the boundary of a certain region $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, we estimate the region's perimeter with the expression

$$P = \sum_{i=1}^{n} \sqrt{((x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2)} \tag{8.2}$$

in which n is the number of points of the boundary.

After applying the above rules, we can detect the regions of the nipples. Fig. 8.5(a) shows the regions that have a number of pixels greater than N_p and that lie inside the boundary of the body (in this study N_p was set to 20). Fig. 8.5(b) shows the regions that do not lie in $region_{up}$ or $region_{lw}$ of the thermograms. This step excludes the regions that lie away from the breasts. Fig. 8.5(c) shows the roundness measure of the regions selected in Fig. 8.5(b). As shown, the three regions have a roundness measure of 0.72, 0.75 and 0.80. The regions that have the two highest values correspond to the nipples.

In Algorithm 2, we present the steps of nipple selection from the given candidates. The algorithm receives the list of nipple candidates and outputs the location of the left and right nipples. To understand the proposed nipple selection algorithm, we summarize and explain the notations used in Algorithm 2 in Table 8.1. The main steps of the algorithm are the following:

- **Steps 4-8.** The algorithm suppresses the regions that lie outside the human body boundary (HBB) and that touch or lie $region_{up}$ or $region_{lw}$ from the list of candidates (N).
- **Steps 9-18.** The algorithm identifies a list containing the left nipple candidates (L_{left}) and then finds the region that has the maximum roundness. However,

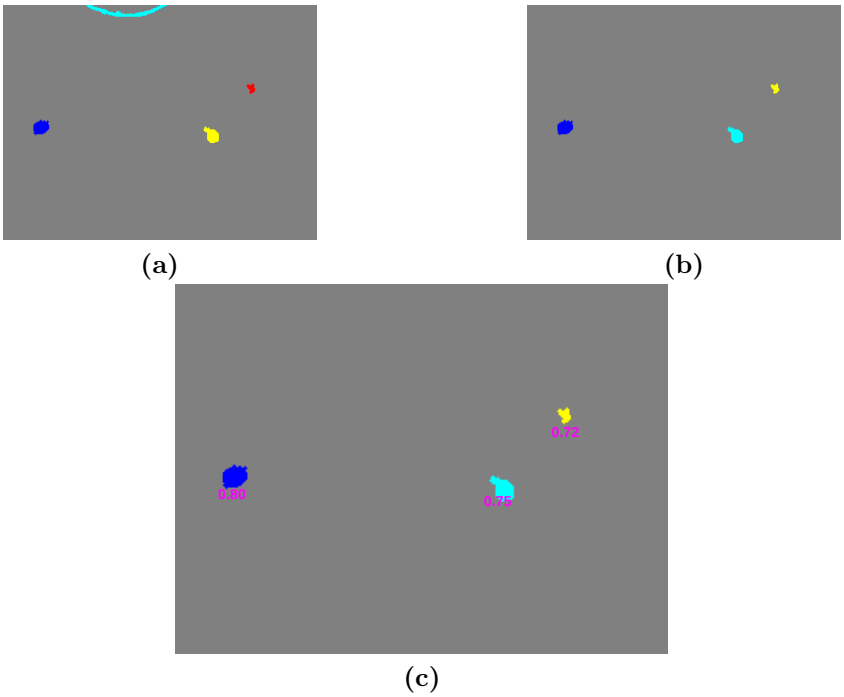


Figure 8.5: Applying the selection rules to detect the nipples. (a) regions having a number of pixels greater than N_p, (b) the regions that lay outside $region_{up}$ and $region_{lw}$, and (c) the roundness of the selected regions

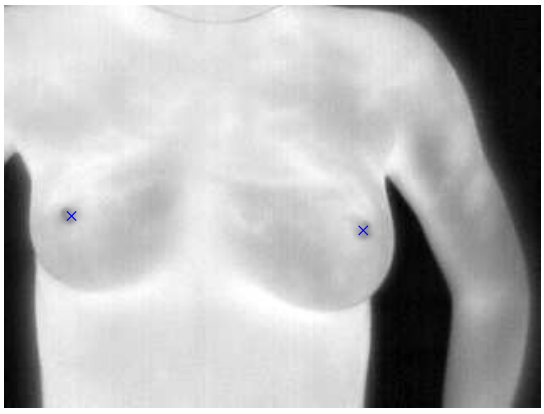


Figure 8.6: Detected nipples

if two regions have the same maximum roundness, it selects the region that has the biggest area. Note that, if L_{left} includes one region, the algorithm directly identifies it as a left nipple.

- **Steps 19-28.** The algorithm repeats steps 9-18 to detect the right nipple.

Finally, we calculate the centroid of each region and plot it on the input thermogram to highlight the location of the detected nipple. Fig. 8.6 shows the final position of the nipples in the input thermogram.

Algorithm 2 Nipple selection algorithm

Inputs: The list of nipple candidates N, HBB and the center-line (L_{cnt})
Outputs: Locations of the two nipples ($nipple_{left}$ and $nipple_{right}$)
procedure NIPPLE SELECTION
    for each candidate $i \in N$ **do**
        Delete i, if any element of $i \notin HBB$
        Delete i, if any element of $i \in region_{up}$
        Delete i, if any element of $i \in region_{lw}$
    end for
    Find left nipple candidates N_{left} that lay at the left of L_{cnt}
    if $length(L_{left}) > 1$ **then**
        $m_l = max\{R(L_{left})\}$
        if $length(m_l) > 1$ **then**
            $m'_l = max\{A(L_{left}(m_l))\}$
            $nipple_{left} = L_{left}(m'_l)$
        else $nipple_{left} = L_{left}(m_l)$
        end if
    else $nipple_{left} = L_{left}$
    end if
    Find right nipple candidates (L_{right}) that lay at the right of (L_{cnt})
    if $length(L_{right}) > 1$ **then**
        $m_r = max\{R(L_{right})\}$
        if $length(m_r) > 1$ **then**
            $m'_r = max\{A(L_{right}(m_r))\}$
            $nipple_{right} = L_{right}(m'_r)$
        else $nipple_{right} = L_{right}(m_r)$
        end if
    else $nipple_{right} = L_{right}$
    end if
end procedure

Table 8.1: Notation used in Algorithm 2

Notation	Meaning
N	List of nipple candidates.
HBB	Human body boundary.
A	The area of a given region.
R	The roundness of a given region.
L_{cnt}	The center-line of the image.
$region_{up}$	Indices of the upper region.
$region_{lw}$	Indices of the lower region.
L_{left}	List of left nipple's candidates.
L_{right}	List of right nipple's candidates.
max	A function that finds the index of the region that has the maximum roundness or biggest area.
m_l	Index of the region that has the maximum roundness in L_{left}.
m_r	Index of the region that has the maximum roundness in L_{right}.
m'_l	Index of the region that has the biggest area in N_{left}.
m'_r	Index of the region that has the biggest area in N_{right}.
$nipple_{left}$	The left nipple.
$nipple_{right}$	The right nipple.

8.3 Experimental results and discussion

The proposed method has been tested using a breast thermography dataset that contains 148 thermograms. We collected the dataset from the Proeng database (Silva et al., 2014). We described this database in Section 2.5.

8.3.1 Evaluation

We used the Recall, Precision and F-score to evaluate the proposed method. If the centroid of the selected region is inside the region of the true nipple, then it is a true positive (TP). On the other hand, if the centroid of the selected region is outside the region of the true nipple, it is a false positive (FP). To calculate the recall, precision and F-score, we determine the following identities:

- TP = correctly identified nipples.
- FP= incorrectly identified nipples.
- P_{pos}=the number of positive instances (number of nipples).

Then, the Recall, Precision and F-score can be calculated as follows:

$$Recall = TP/P_{pos} \tag{8.3}$$

$$Precision = TP/(TP + FP) \tag{8.4}$$

$$\text{F-score} = \frac{2.precision.recall}{(precision + recall)} \tag{8.5}$$

8.3.2 Results

Fig. 8.7 shows examples of correctly detected nipples on thermograms for different women. We show the results of the proposed method with small breasts (Fig. 8.7 (c,d,g)), medium breasts (Fig. 8.7 (e,h,i)) and big breasts (Fig. 8.7(a,b,f,j−o)). As shown, in all cases the proposed method correctly determines the location of the nipples.

We evaluated the proposed method with the whole dataset. Note that the thermograms used contain at least one nipple in their profile. Indeed, we have excluded the thermograms that do not include any nipples in their profiles. However, the dataset still contains very difficult cases, where nipple detection is difficult even for humans. Table 8.2 shows the evaluation of the proposed method with the whole dataset. The nipple detection results were outstanding (precision close to 0.99). Note that the dataset used includes several cases in which it is difficult to identify the nipples even with the human eye.

Table 8.2: Evaluation of the performance of the proposed method

Method	**Recall**	**Precision**	**F-score**	**Time (s)**
Proposed	**0.930**	**0.989**	**0.958**	**0.30**
(Koay et al., 2004)	0.500	0.652	0.566	0.91

Fig. 8.8 shows examples of false positive nipples. As shown, the algorithm fails to detect some nipples in these cases. Indeed, the regions of undetected nipples are not clear in the images, so detecting the nipples with the human eye is also difficult.

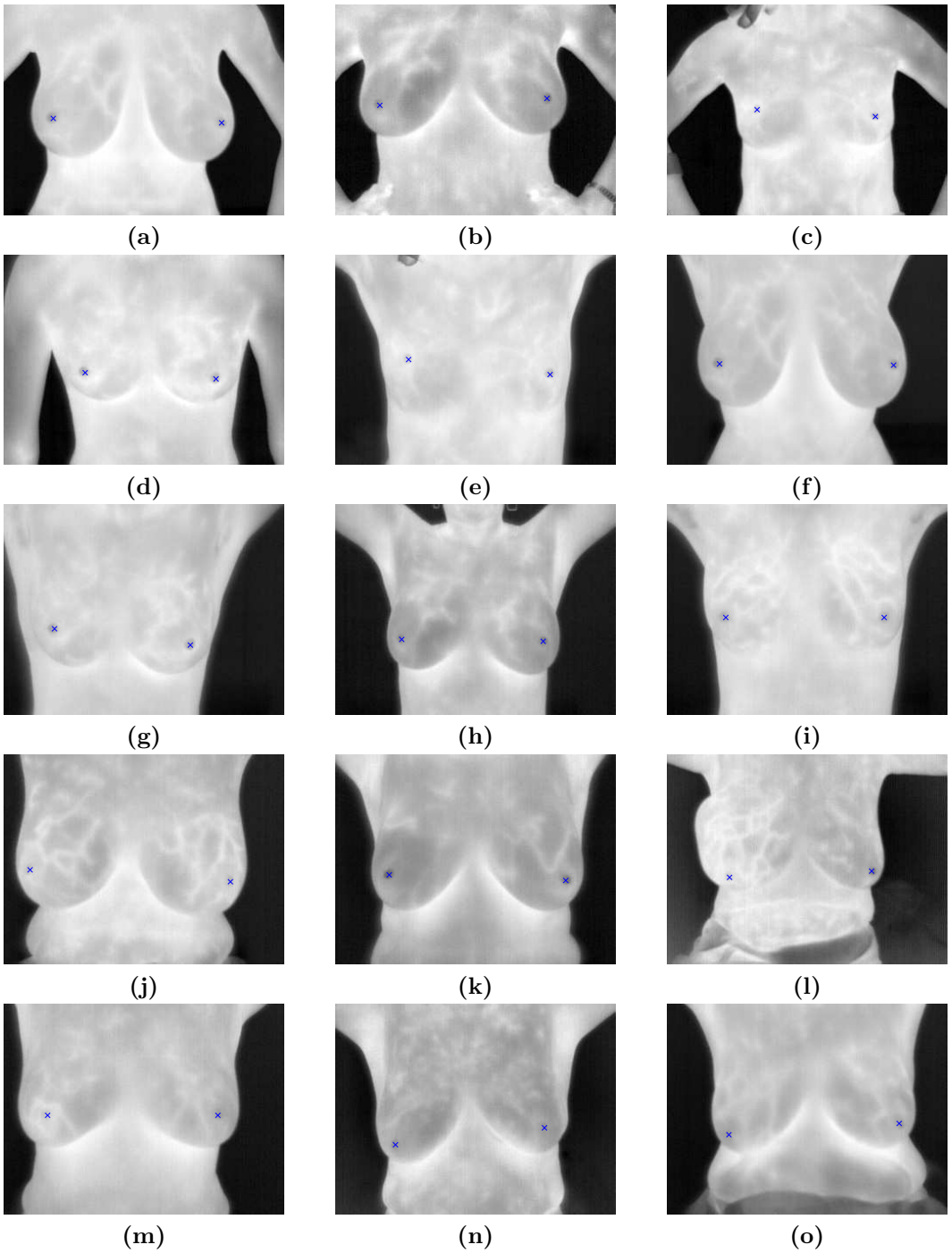
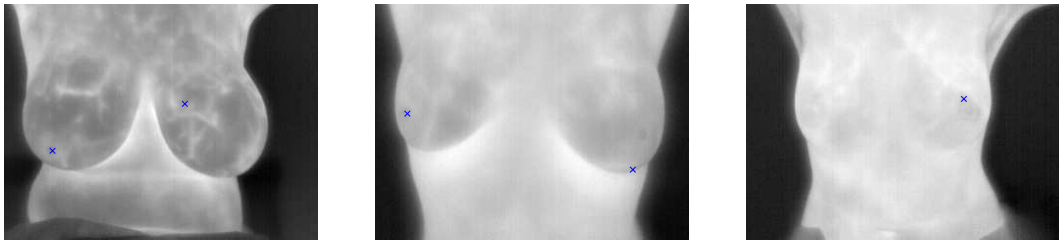


Figure 8.7: Examples of correctly detected nipples using the proposed method

Consequently, the nipples can not be detected using the proposed method in these cases.



(a) (b) (c)

Figure 8.8: Examples of false positives

8.3.3 Comparison with related work

Indeed, very few methods have been proposed to detect nipples in thermograms. In this subsection we compare our method with the approach suggested in (Koay et al., 2004), in which the authors used an adaptive thresholding and morphological operations to detect the nipples. They also removed big contour regions and regions in the upper breasts, and assumed that the nipples have an eccentricity close to one. Indeed, they did not present any parameter settings and made no evaluation of nipple detection because the main goal of their work was to classify thermograms into normal, fibrocystic or cancerous. We implemented the aforementioned rules and tuned the values of the parameters to obtain the best performance.

As can be seen in Table 8.2, the proposed method easily outperforms the method of (Koay et al., 2004) in terms of recall, precision and F-score. Fig. 8.9 (a-b) shows the performance of the two methods in a thermogram containing two nipples. The proposed method easily detects the two nipples in the cases presented (Fig. 8.9 (a)) while (Koay et al., 2004) does not (Fig. 8.9 (b)). Fig. 8.9 (c-d) shows the performance of the two methods in a thermogram containing one clear nipple (right nipple) in its profile (the left nipple is fuzzy). The proposed method successfully detects the right nipple (Fig. 8.9 (c)) while (Koay et al., 2004) does not (Fig. 8.9 (d)). These results demonstrate that the rules used in (Koay et al., 2004) produce poor detection results.

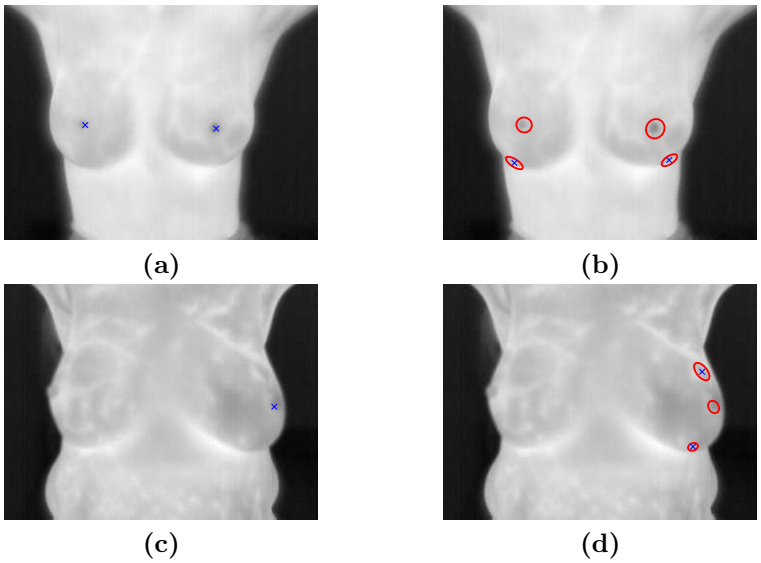


Figure 8.9: Comparing the performance of proposed method (a,c) and the approach proposed in (Koay et al., 2004) (b,d)

The two methods were implemented using MATLAB on an Intel processor core 2 Quad at 2.5GHz and 8 GB of RAM. As presented in Table 8.2, the proposed method takes 0.3s to determine the nipples. This time is approximately one-third of the execution time of (Koay et al., 2004), indicating that the proposed method is more suitable for working with real-time applications.

8.3.4 Possible applications of the proposed method

As we have shown above, our method accurately detects the nipples in real-time (0.3s). Thus, it could be used in such applications as the following:

- It can be used to find two seed points to segment the breast region into thermograms. Thus, we can integrate our method with the cancer detection approach proposed in (Schaefer et al., 2009) where the breasts were manually segmented. We suggest replacing this step by an automatic segmentation method based on the method suggested in this study. This makes the approach proposed in (Schaefer et al., 2009) fully automatic.
- Given two thermograms, the proposed method can be used to register one breast to the coordinates of another. Thus, it can be used to improve the registration step used in (Saniei et al., 2015).

- It can be used to find the correspondence location of masses in images that were acquired using different modalities (e.g., thermograms and mammograms) for the same breast.
- It can be used to design a robust thermogram registration framework. In chapter 9 we used an optimization approach to find the optimal location of the origin of the curvilinear coordinates to register mammograms; in turn, the nipples detected by our method will be used as origins to establish the curvilinear coordinates in order to align thermograms.

8.3.5 Limitations

Like any computer approach, the proposed method has some limitations:

- It does not work in thermograms in which nipples do not appear or are not clearly visible.
- It fails to detect nipples in thermograms that contain fuzzy regions, and in which even the human eye would have trouble detecting them.

8.4 Conclusion

An unsupervised, automatic, accurate, simple and real-time method for detecting nipples in thermograms is proposed in this chapter. The main stages of the proposed method are: human body segmentation, determination of nipple candidates using adaptive thresholding and finally detection of the nipples using rules derived from the anatomical structure of the human body. The proposed method takes 0.3s to detect the nipples. It gives very accurate results in real-time.

In the next section we propose a registration method for aligning temporal mammograms. We also propose a method for quantifying and visualizing the changes in breast tumors in patients undergoing medical treatment through strain tensors.

Part V

Analysis of temporal evolution of breast cancer

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

CHAPTER 9

Temporal mammogram image registration using optimized curvilinear coordinates

9.1 Introduction

The goal of *image registration* is to find the optimal transformation function that aligns one image with another. In other words, the registration process should bring the coordinates of the current (template) mammogram to the coordinates of the previous one. Registration helps doctors analyze and visualize mammograms. The comparison of mammograms requires a registration (alignment) method. Four breast mammogram comparisons are usually performed (Oliver et al., 2010):

- *Temporal analysis* is performed between mammograms for the same breast at different screening examinations.

- *Bilateral analysis* is performed between mammograms of the left and right breasts of the same woman.
- *Ipsilateral analysis* is performed between the CC and MLO views of the same breast.
- Analysis of breast images acquired from different *modalities* (e.g. X-ray with MRI).

A full review of medical image registration methods can be found in (Oliveira and Tavares, 2014; Sotiras et al., 2013). Few studies have evaluated the performance of image registration methods using mammograms. In (van Engeland et al., 2003), four registration methods with MLO-MLO or CC-CC mammogram pairs were evaluated. The methods used were: nipple-based alignment, center of mass alignment, warping based on manually selected control points and MI registration method. The MI-based method yielded the best registration results, whereas, the control points-based method performed worst. Diez et al. (2011) presented a quantitative evaluation of state-of-the-art intensity-based image registration methods applied to mammographic images. The study assessed the suitability of global rigid transformation and local deformable registration methods for mammographic image analysis. They showed that local deformations (multi-resolution B-Spline deformations) provide the most accurate registration results. In (Pereira et al., 2010), a comparison between affine, fluid and free-form deformation registration methods was shown. The study concluded that the affine method gave the best registration results. The main drawbacks of the existing mammogram registration methods are that they use either non-realistic global transformations or local deformable models. The global registration methods cannot properly cope with the local deformations; in turn, the local deformation models may yield unrealistic deformations.

In this chapter, we propose a framework for temporal mammogram registration based on a transformation model derived from the breast anatomy: namely the *curvilinear coordinate system*. In curvilinear mapping, a coordinate pair (s, t) is assigned to each pixel in Cartesian coordinates, (x, y). The theoretical point of view of the proposed representation was inspired from the work proposed in (Daugman,

2004) where a dimensionless polar coordinate system was used to generate a new representation for iris images. The construction of the curvilinear coordinates does not require any information about the internal structures of the breast. To construct the curvilinear coordinates, we use the breast boundary and a reference point located on it. Thus the resulted representation of a given mammogram is invariant to changes in the size, position and orientation of the internal structures of the breast (the internal structures do not play any role on the construction process). We utilized the curvilinear coordinates to cope both with global and local deformations in the breast area and compensate the deformations between the mammograms.

Fig. 9.1 shows a mammogram that has been forwardly transformed from the Cartesian coordinates to the curvilinear coordinates, and then it has been inversely transformed from the curvilinear coordinates to the Cartesian coordinates. The use of curvilinear coordinates in mammogram registration enables us to build a registration approach based on a reasonable grid which mimics the anatomy of the breast, instead of using the Cartesian grid which is composed of vertical and horizontal lines. In addition, mammogram registration based on the curvilinear coordinates does not require control points to be determined nor a correspondence algorithm to be used. The parameters of curvilinear coordinates may be optimized, as will be shown later, to find the best alignment between the reference and template images.

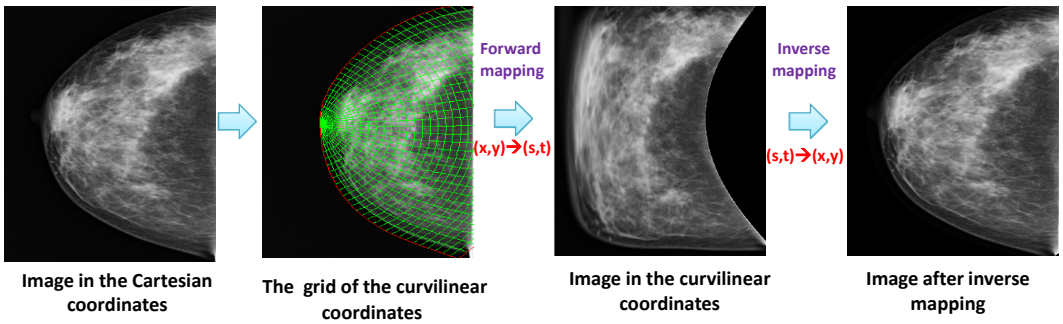


Figure 9.1: The curvilinear coordinates

9.2 Related work

In general, image registration methods can be divided into *feature-based* and *intensity-based* methods. However, with mammogram registration, there are more specific registration categories such as *breast contour-based*, *anatomical structures-based* and *image representation-based* methods. Here, we discuss the the work done on each category.

9.2.1 Feature-based registration methods

Feature-based methods depend on the features extracted from mammograms. Control points are the most commonly used features for registrating mammograms. Marias et al. (1999) used a set of breast boundary points and internal points with thin plate spline to register a pair of temporal mammograms. They used a matching approach to detect the potential internal points by comparing the features of the local maximum and minimum points in the breast region in the reference and template mammograms. Hong and Brady (2005) also proposed a method for comparing mammogram pairs. They extracted salient regions in a topological way. An integral invariant representation of shape, in combination with area and distance measures, was used to establish the correspondences between mammograms. Furthermore, Wai and Brady (2005) proposed a curvilinear structure-based mammogram registration approach in which they incorporated junctions of curvilinear structures as internal landmarks. The curvilinear structures describe connective tissue, blood vessels and mammary ducts, which are detected by an algorithm based on the monogenic signal. The junctions are extracted using a local energy-based method, which considers the orientation information provided by the monogenic signal.

Indeed, it is difficult to extract consistent features from the mammograms because the appearance of the breast region depends heavily on the strength of the compression which was applied to the breast using the compression paddles; therefore, few distinct landmarks can be determined. The main drawback of the aforementioned methods is that the extracted features (e.g. points and regions) may differ due to

the compression, and this may cause inaccurate registration results.

9.2.2 Intensity-based registration methods

Intensity-based registration depends on the image pixel values. The illumination change between mammograms and the mapping model affects the accuracy of these methods. Richard and Cohen (2003) combined a region matching procedure and segmentation by formulating the energy minimization problem with free boundary conditions. A multi-grid implementation with a coarse-to-fine strategy was used to reduce the execution time. The gradient descent algorithm was used to optimize the energy. The main drawback of this method is that the gradient descent may get trapped into a local minimum, yielding inaccurate registration results; in addition, the accuracy of the segmentation process controls the success of the overall registration model.

In (Rueckert et al., 1999) an affine transformation was used to recover the global deformations between magnetic resonance breast images followed by a free-form deformation based on B-splines to model the local deformations. Normalized mutual information was used as a similarity measure. Rohlfing et al. (2003) added a regularization term to (Rueckert et al., 1999) to preserve the local volumes. Fischer and Modersitzki (2004) proposed a unified approach to fast image registration and a new curvature-based registration technique. The internal forces are designed to minimize the curvature of the displacement field. The illumination change between mammograms and the mapping model affects the accuracy of these methods.

9.2.3 Breast contour-based registration methods

Breast contour alignment provides a good basis for registering the mammograms. Sallam and Bowyer (1999) proposed an automated technique for identifying differences between corresponding mammograms. Their method recovered an approximate deformation between a pair of mammograms based on the determination of corresponding features across the two mammograms. The registration process was completed by using an un-warping technique to transform

the template mammogram into the coordinate system of the reference mammogram. Marias et al. (2000) proposed a multi-scale landmark selection approach to improve the accuracy of temporal mammogram registration. They explained the need to establish correspondences between internal regions in temporal mammograms for robust and more accurate registration results. On the basis of automatically detected boundary landmarks, they made a partial registration and subsequently analyzed the mammogram pair using a non-linear wavelet scale-space to identify the significant regions of interest. They showed that a small (but significant) number of internal correspondences greatly improved the registration accuracy and better approximated the complex internal tissue deformation in the breast region. In addition, in (Marias et al., 2005) they proposed a three-step framework for the registration of mammogram sequences: boundary registration, extraction of the correspondence between internal points using wavelet-based analysis, and, finally, thin plate spline transformation to un-warp the template mammogram using the boundary and the internal points. Although the registration results were acceptable, the accuracy of these methods depends mainly on the internal points. The compression applied during the acquisition process deforms the internal regions of the breast, which may yield an error in the determination of the internal control points. These errors may produce bad registration results.

9.2.4 Anatomical structures-based registration methods

Some studies have used the anatomical structures of the breast in mammogram registration. Rajagopal et al. (2008) used an anatomically realistic bio-mechanical model of the breast. A model was customized to each breast. These models were generated by fitting a geometrical model to segmented data from breast MRIs. Boucher et al. (2010) proposed a two-step mammogram registration approach. In the first step, they used the pectoral muscle and the nipple to make a rigid alignment of MLO mammogram pairs. This step minimizes the differences caused by positioning and compression. In the second step, some deformation based on linear scalings was applied to local regions in the mammograms. This method can not be used to

register CC mammographic pairs.

9.2.5 Image representation-based registration methods

Image representations may improve the registration accuracy. Image representations should be invariant to changes in the size, position and orientation of internal structures. In (Daugman, 1993, 2004), a representation for iris images was proposed. This representation used a dimensionless polar coordinate system to generate a new representation for each iris image. The invariance to the aforementioned factors was achieved by mapping the iris images onto a double dimensionless pseudo-polar coordinate system. In (Daugman, 1993), the rubber sheet model was used to assign each pixel of the iris image, regardless of its size and papillary dilation, with a pair of real coordinates (r,θ), where r is in the unit interval $[0, 1]$ and θ is an angle in [0,2π].

The above methodology prompted us to build a new coordinate system for breast mammograms (the curvilinear coordinates), in which the breast boundary and the nipple can be used as a basis for establishing these coordinates for each mammogram. Like Daugman's coordinates (Daugman, 1993), the curvilinear coordinates generated for each mammogram are invariant to changes in the size, position and orientation of the internal structures.

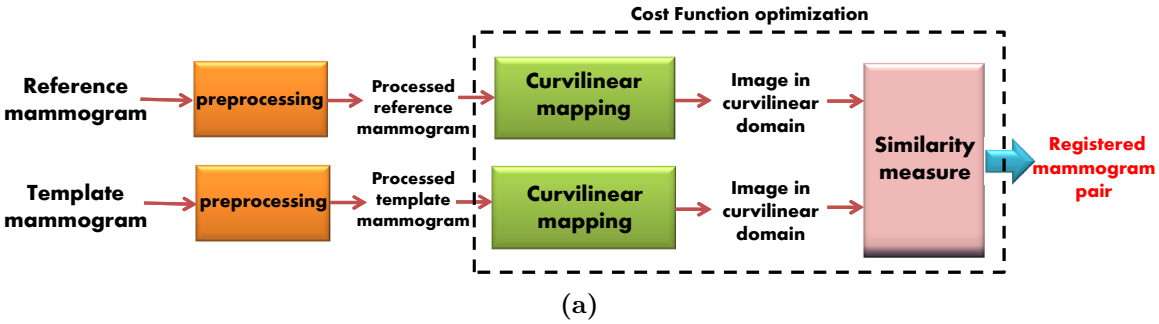
Brandt et al. (2011) used anatomical coordinates to represent the breast region within a mammogram. These coordinates were based on the breast anatomy of 2D medio-lateral (ML) or MLO view mammograms. The mammograms were registered according to the location of the pectoral muscle, the nipple and the shape of the breast boundary. A second-order approximation for the breast boundary was based on the location of the nipple, two points in the breast boundary and the pectoral muscle. Then, a non-linear mapping was used to map one image to the coordinates of the second image. Despite the effectiveness of the method in (Brandt et al., 2011), it can not be applied to register CC-CC or CC-MLO mammogram pairs because CC mammograms do not contain pectoral muscle. The manual selection of these points is the main disadvantage of this method because the construction

of the coordinates (which map the image) is sensitive to the accuracy of the user. Moreover, second-order polynomials are insufficient to approximate the breast boundary under different compression levels. In addition, detecting the location of the nipple is particularly difficult in mammograms which do not show the nipple in the mammogram profile. Unlike Brandt's method (Brandt et al., 2011), the proposed method is fully automatic (no need for manual inputs) and it can also be applied to both CC-CC and MLO-MLO mammographic pairs.

9.3 Methods

A dataset of 100 temporal mammogram pairs is used in our experiments. It was collected from the Hospital Universitari Sant Joan (Reus, Spain). The temporal mammograms are *full field digital mammograms (FFDM)* with sizes of 1770×2370, 3540×4740 and 2364×2964 pixels, and with spatial resolution of 100μ. The image size depends on the compression plates used in the acquisition process (the amount of the compression varies according to the breast size of each woman). The images collected were saved in the digital imaging and communication in medicine (DICOM) format.

Prior to registration, each mammogram was preprocessed to extract the breast area, identify the orientation and determine the breast boundary. The curvilinear coordinates are used to represent the breast region in each mammogram. The breast region representation using the curvilinear coordinates is based on the breast boundary and an arbitrary reference point (origin). The breast boundary is used as a centreline, and to generate the curvilinear grid. An exhaustive search technique is used to find the optimal reference points which maximize the MI between the reference and template mammograms. The proposed method can properly align both the breast boundary and the internal structures of the breast region. Fig. 9.2 presents the main steps of the proposed registration approach.



(a)

Figure 9.2: Mammogram registration framework

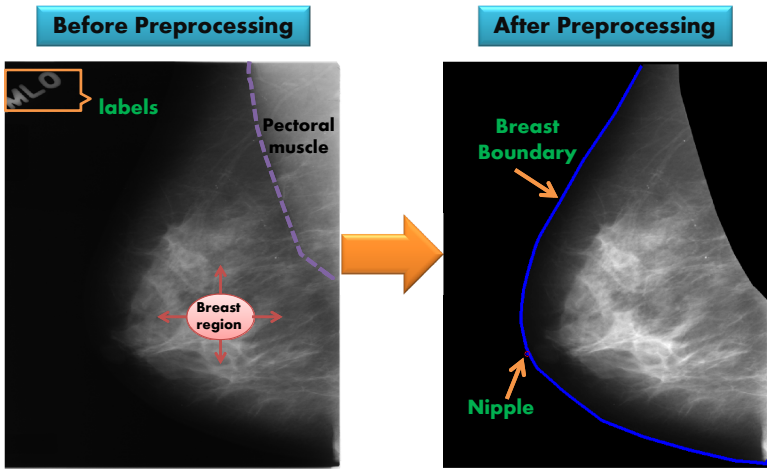
9.3.1 Preprocessing stage

The steps of the preprocessing stage can be summarized as follows:

1. The mammograms used in this study were of different sizes, so they were resized into a common size. We tried many image sizes and found that 1024×1024 pixels maintains the performance of the proposed method with little computational complexity.
2. If accuracy is to be acceptable, the mammograms need to be enhanced without distorting the soft details. We used the median filter for noise suppression and image enhancement. The median filter is used to generate a low-frequency image by replacing the pixel value of each pixel in the mammogram with the median pixel value computed over a square area of 11×11 pixels centered at the pixel location (Sun and Neuvo, 1994).
3. The breast region is extracted from the mammogram image by removing the artifacts and background. To do this, we used the approach proposed by Subashini et al. (2010). The intensity values of mammograms were scaled to the range $[0, 255]$ and a threshold of 18 was used to convert each mammogram to a binary image (mask). The connected component labeling algorithm was used to select the largest object from each binary image (the breast area and the pectoral muscle). Then, morphological operations were used to remove the isolated pixels (clean operation) and reduce noise (majority operation, 3×3 neighborhood). We used the resulting mask to remove the labels and the artifacts from the mammograms.

4. The mammogram has the image of the breast at the center with *left* or *right* orientation. In the first case, the chest wall is to the left of the mammogram image, whereas in the second, it is to the right. For mammograms to be registered, they should have the same orientation. In (Chandrasekhar and Attikiouzel, 1996), pixels near the left and right corners were extracted and the pixel with brightest value determined the orientation of the breast. Given the binary mask resulting from step 3, we removed 1/8 of the number of rows from the top and bottom borders and 1/8 of the number of columns from the left and the right sides; this reduces the black boundaries surrounding the breast region. Then the sum of the first column (S_f) and the sum of the last column (S_l) were calculated. If $S_f < S_l$, the breast is right-oriented; otherwise, it is left-oriented. In this study, we work with right-oriented images. After the orientation has been estimated, all mammograms are converted to the same orientation.
5. To eliminate the pectoral muscle, we used the binary mask resulting from step 3 and the mammogram orientation which we determined in step 4 to determine the location of the seed of a region growing segmentation (threshold = 30). If the mammogram was left-oriented, the location of the seed was set to (5^{th} row , 5^{th} column), and if it was right-oriented, the location of the seed was set to (5^{th} row, *number of columns-5*). After these operations, the mask of the pectoral muscle was generated. To smooth this mask, we used erosion and dilation morphological operations with a flat, disk-shaped structuring element (the radius was set to 5).
6. To detect the breast skin boundary, we used the algorithm proposed by Maitra et al. (2011). This algorithm consists of two stages: image enhancement and edge detection. The breast contour is extracted by defining the outermost edges, so the starting point of the scanning process should be determined. Each row of the mammogram is scanned from the left to determine the right most edge points. This scanning process is repeated for all the rows to find the breast contour.

Fig. 9.3 shows an example for a mammogram image before and after the preprocessing step. We can notice the absence of the pectoral muscle and the labels in the processed image; in addition, the estimated breast boundary is shown in blue colour.



(a)

Figure 9.3: Result of the preprocessing stage

9.3.2 Curvilinear mapping

In curvilinear mapping, a coordinate pair (s, t) is assigned to each pixel in Cartesian coordinates, (x, y). This operation represents a direct mapping from Cartesian to curvilinear coordinates

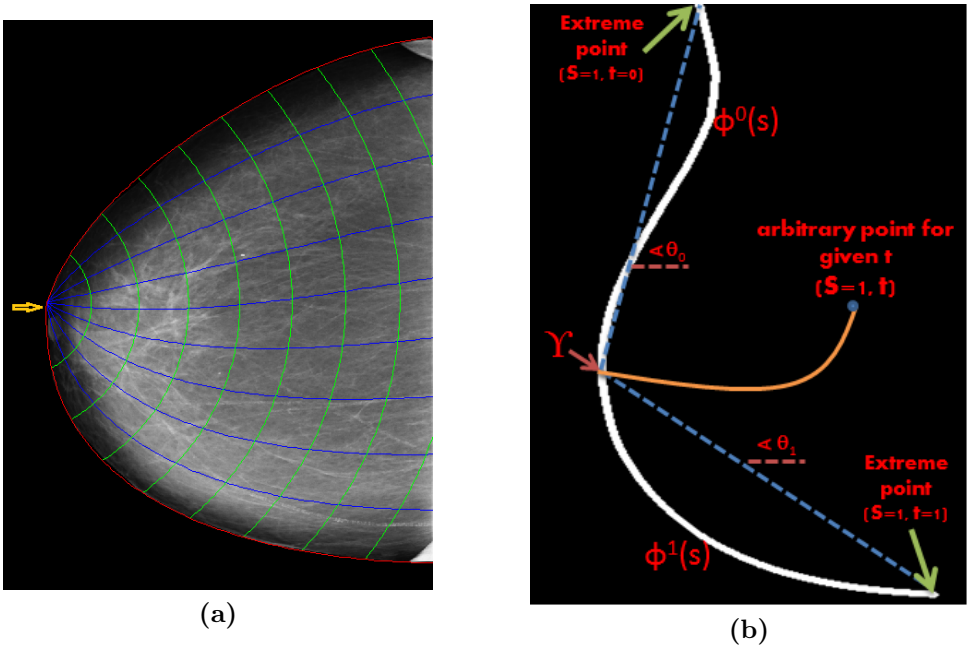
$$(x, y) \xrightarrow{M_c} (s, t) \tag{9.1}$$

where c is the breast skin boundary that generates the curvilinear coordinates and M_c is the mapping operator which transforms the image from Cartesian coordinates $I(x, y)$ to curvilinear coordinates $I(s, t)$ as follows:

$$I(s, t) = M_c \{I(x, y)\} \tag{9.2}$$

In curvilinear mapping, each pixel is identified by its *surface distance*, s, and *penetration distance*, t. The surface distance, s, is measured from the reference point of the coordinate system (point γ on the breast skin boundary) while t is

measured from the breast's surface. The direct mapping M_c can be interpreted as a boundary-driven function that computes the (s, t) coordinates that correspond to each (x, y) Cartesian coordinate pair.



(a) (b)

Figure 9.4: Curvilinear grid, (a) generation of the curvilinear grid, (b) explanation of the curvilinear coordinates

As shown in Fig. 9.4(a), the process starts with the breast skin boundary c (red curve) and reference point γ (this point belongs to the breast boundary c, see the yellow arrow in Fig. 9.4(a)). Then s-coordinate (green curves) and t-coordinate (blue curves) are generated. The domain of the s-coordinate is normalized to $s \in [0, 1]$, in correspondence with the two extreme points of the breast boundary around the reference point. The domain of the t-coordinate is also normalized to $t \in [0, 1]$.

The curvilinear coordinates can be generated from the breast boundary c as follows. Let the breast boundary be $c = (x(u), y(u))$ with length L. The surface distance s of a point (x, y) is defined as:

$$s(x, y) = \left| \frac{2}{L} \int \sqrt{\left(\frac{dx}{du}\right)^2 + \left(\frac{dy}{du}\right)^2} du - 1 \right| \tag{9.3}$$

As illustrated in Fig. 9.4 (b), the two segments around the reference point (γ) can be approximated using two N-degree parametric polynomials of s, $\phi^0(s)$ and $\phi^1(s)$:

$$\begin{aligned} \phi^1(s) &= (x^1, y^1), \\ \phi^0(s) &= (x^0, y^0) \end{aligned} \tag{9.4}$$

These polynomials correspond to the parametric curves for all the points in curvilinear coordinates with $t = 0$ and $t = 1$, respectively. The variables x^0, y^0, x^1 and y^1 can be computed as follows.

$$\begin{aligned} x^0 &= \sum_{n=1}^{N} P_{0,n} s^n, \\ y^0 &= \sum_{n=1}^{N} P_{1,n} s^n, \\ x^1 &= \sum_{n=1}^{N} P_{2,n} s^n, \\ y^1 &= \sum_{n=1}^{N} P_{3,n} s^n \end{aligned} \tag{9.5}$$

where $P_{0,n}$, $P_{1,n}$, $P_{2,n}$ and $P_{3,n}$ are constant coefficients. In this study N is set to 4. To generate the parametric curves for all the points at any t (the orange curve in Fig. 9.4(b)), we define the auxiliary curve $\bar{\phi}(\bar{x}, \bar{y})$ as follows.

$$\begin{aligned} \bar{x} &= (t-1)(x^0 cos\theta_0 + y^0 sin\theta_0) + t(x^1 cos\theta_1 + y^1 sin\theta_1), \\ \bar{y} &= (t-1)(y^0 cos\theta_0 - x^0 sin\theta_0) + t(y^1 cos\theta_1 - x^1 sin\theta_1) \end{aligned} \tag{9.6}$$

where θ_0 and θ_1 are the angles at the extreme points of $\phi^0(s)$ and $\phi^1(s)$ (see Fig. 9.4(b)), by substituting $s = 1$ in Eq. 9.5, they are calculated as follows.

$$\begin{aligned} \theta_0 &= \tan^{-1}(\tfrac{y^0}{x^0}), \\ \theta_1 &= \tan^{-1}(\tfrac{y^1}{x^1}) \end{aligned} \tag{9.7}$$

The analytical expression that transforms each (x, y) pair into its corresponding (s, t) pair is calculated by rotating the auxiliary curve in Eq. 9.6 by $\theta(t)$,

$$\begin{aligned} x(s,t) &= \bar{x}cos\theta(t) - \bar{y}sin\theta(t), \\ y(s,t) &= \bar{x}sin\theta(t) + \bar{y}cos\theta(t) \end{aligned} \tag{9.8}$$

The angle $\theta(t) = (\theta_1 - \theta_0)t + \theta_0$ is a function of t and the extreme angles θ_0 and θ_1, and it represents the rotation angle that goes from θ_0 to θ_1 as t moves from 0 to 1. Fig. 9.5 shows examples of the curvilinear mapping. For each mammogram we show the breast boundary, the curvilinear grid and the (s, t) mapping. For the sake of clear display for the curvilinear grid, few points of s- and t-coordinates are plotted. As shown, the two parametric polynomials, $\phi^0(s)$, and $\phi^1(s)$, properly fit the breast boundary (they yielded an interpolation error lower than 0.5%); moreover, they do not stretch inside breast region. As we mentioned before, N is set to 4. This value yields a good approximation for breast contours (see the examples of Fig. 9.5).

9.3.3 Inverse mapping and registration

As the curvilinear mapping M_c establishs a one-to-one correspondence between each coordinate pair $\left[(x, y) \xrightarrow{M_c} (s, t)\right]$, it is possible to invert the curvilinear mapping to recover the original image in Cartesian space as follows:

$$I(x, y) = M_c^{-1}\{I(s, t)\} \tag{9.9}$$

In other words, the inverse mapping M_c^{-1} transforms each curvilinear coordinate pair (s, t) to its corresponding Cartesian coordinate pair (x, y) as follows:

$$(x, y) \xleftarrow{M_c^{-1}} (s, t) \tag{9.10}$$

To align the template mammogram I_t with the coordinates of the reference mammogram I_r, they are both mapped to the curvilinear coordinate system as

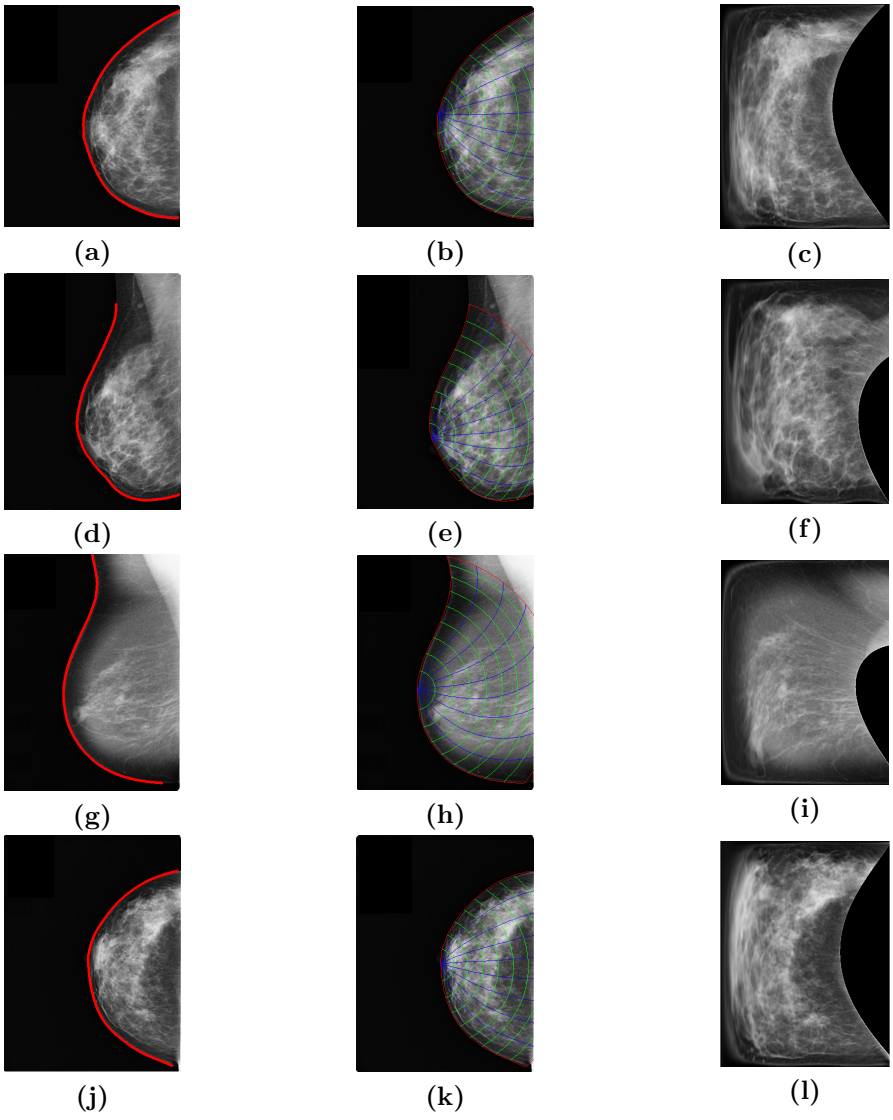


Figure 9.5: Examples of curvilinear mapping. The left column shows the breast boundary of each mammogram. The middle column shows the curvilinear grid which is generated from the breast boundary. The right column shows the (s, t) mapping of each mammogram

follows:

$$\begin{aligned} I_r(s,t) &= M_{c_r}\left\{I_r(x,y)\right\}, \\ I_t(s,t) &= M_{c_t}\left\{I_t(x,y)\right\} \end{aligned} \tag{9.11}$$

where $I_r(s,t)$ is the curvilinear mapping of $I_r(x,y)$ using the breast boundary of the reference mammogram C_r, and $I_t(s,t)$ is the curvilinear mapping of $I_t(x,y)$ using the breast boundary of the template mammogram C_t. The registered mammogram I_{reg} can be estimated by reverse mapping the template mammogram using the boundary of reference mammogram C_r as follows:

$$I_{reg}(x,y) \xleftarrow{M_{c_r}^{-1}} I_t(s,t) \tag{9.12}$$

where $M_{c_r}^{-1}$ is the reverse mapping using the boundary of the reference mammogram C_r.

Since Eq. 9.8 cannot be analytically inverted, the inverse mapping M_c^{-1} can only be obtained numerically. We used the *Quickhull* algorithm (Barber et al., 1996) to perform the inverse mapping (it yielded an interpolation error lower than 0.5%). Note that both the direct and inverse mapping can be implemented in one step. That is, it is possible to implement the operation $M_{c_r}^{-1}\{M_{c_t}\{I_t(x,y)\}\}$ without needing to perform inverse mapping operations.

The position of the reference point (origin) of the curvilinear grid may be assumed to be in the middle of the breast boundary as in (Pertuz et al., 2014). A more accurate assumption is to consider the nipple position as a reference point for the coordinate system. However, the nipple is not visible in most mammograms and it also appears as a region in the mammogram, so choosing a reference point from that region is problematic. Fig. 9.6 shows the curvilinear grid at six different positions (P1-P6) for the reference point (the yellow arrow points to the reference point of the grid). It is clear that P4 is the most suitable position at which to establish the curvilinear grid.

Fig. 9.7 shows the sensitivity of the registration process to the position of the reference point of the curvilinear mapping. As the position of the reference point changes, the MI between the registered and the reference mammograms changes.

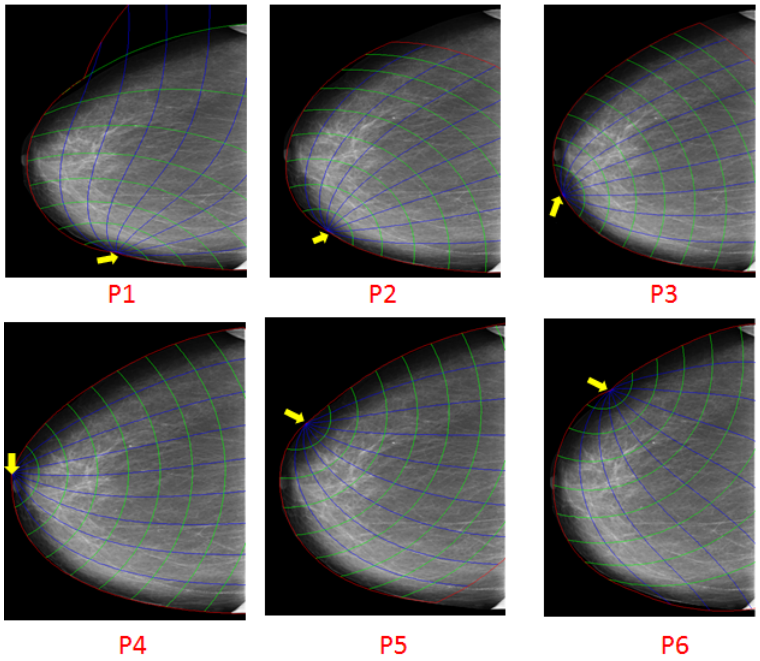


Figure 9.6: The curvilinear grid at six different positions (P1-P6) of the reference point

The indices of the breast boundary are normalized to $[0, 1]$, and the position of the reference point is set at one of the breast boundary indices. In Fig. 9.7, the *x-axis* represents the possible reference point positions of the reference mammogram while the *y-axis* represents the possible reference point positions of the template mammogram. The position which produces the maximum MI will be selected to establish the curvilinear mapping and to estimate the registered mammogram I_{reg}.

Taking the position of the reference point of the reference and template

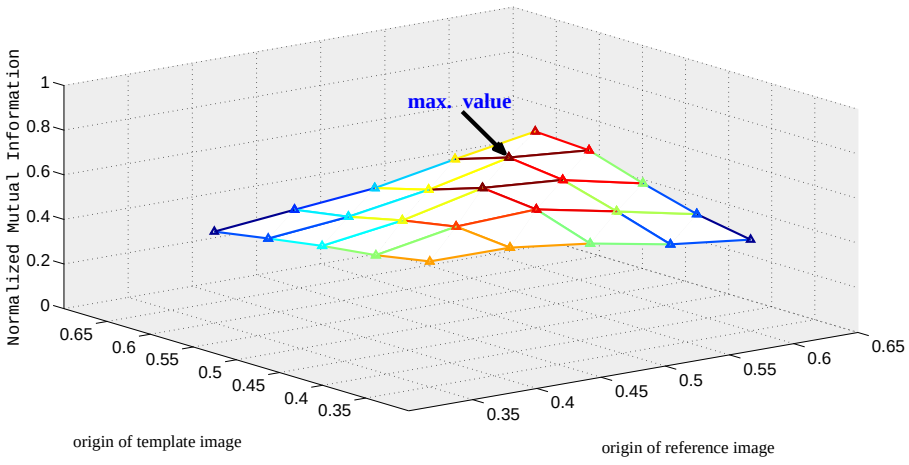


Figure 9.7: Effect of reference point selection on the registration process

mammograms $\gamma = [\gamma_1, \gamma_2]$ into account, the registration process can be formulated as an optimization problem as follows.

$$\gamma^* = \arg\max_{\gamma} S(I_r(s,t), I_t(s,t)) \tag{9.13}$$

where S is a function that measures the degree of similarity between the reference and template mammograms (I_r,I_t) and $\gamma^* = [\gamma_1, \gamma_2]$ is the vector that contains the position of the reference point of the reference mammogram γ_1 and the reference point of the template mammogram γ_2. The *sum of square differences* is often used as a similarity measure in mono-modal image registration because the grey levels are approximately the same in the reference and template images (Guo et al., 2006). However, breast compression and the presence of pathological subjects in mammograms such as lesions invalidate the assumption of grey level constancy between images. For this reason, we used the MI as a similarity measure (cost function). MI measures the dependency between the intensity distributions of two images (Maes et al., 1997). It can work properly with pixels that are non-linearly related.

As there are only two parameters to be optimized (the location of the reference points), we used an exhaustive search procedure to find the optimal values (γ_1, γ_2) which maximize the MI between the reference and registered mammograms. In this study, we restricted the domain of (γ_1, γ_2) to [0.3-0.7] in steps of 0.04. The optimal reference points are fed into the curvilinear mapping to transform the template mammogram to the coordinates of the reference mammogram.

9.4 Experimental results and discussion

To evaluate the ability of the proposed method to register temporal mammograms, we have carried out several experiments. The results are summarized in this section.

9.4.1 Evaluation

Measuring the performance of mammogram registration methods is a particularly difficult task because the deformation may be different in each mammogram and the breast region structures may change due to the growth of tissues or masses. In this study, the *structural similarity* (SSIM) index (Wang et al., 2004) is used to evaluate the proposed registration framework. The SSIM index is an implementation of the idea of structural similarity between two images considering contrast, luminance and structure. An SSIM value of 1 indicates perfect similarity between the two mammograms while 0 indicates that they are completely different. The SSIM index between images x and y is defined as follows.

$$\begin{aligned} SSIM(x,y) &= \frac{(2\mu_x\mu_y+C1)(2\sigma_{xy}+C2)}{(\mu_x^2+\mu_y^2+C1)(\sigma_x^2+\sigma_y^2+C2)}, \\ \sigma_{xy} &= \frac{1}{T-1}\sum_{i=1}^{T}(x_i-\mu_x)(y_i-\mu_y) \end{aligned} \tag{9.14}$$

where T is the total number of pixels in each image, μ_x is the average of x, μ_y is the average of y, σ_x is the standard deviation of x, and σ_y is the standard deviation of y. $C1 = (k_1L)^2$ and $C2 = (k_2L)^2$ are two variables to stabilize the division with small denominator. L is the dynamic range of the pixel values (in our case it is 256). The values of k1 and k2 are 0.01, and 0.03, respectively (Wang et al., 2004).

The *joint histogram* between the reference and the registered mammograms is a visualization of the correspondence between the pixels in the mammograms. As the amount of dispersion in the joint entropy image decreases, the MI between the reference and the registered mammograms increases, indicating that a better geometric alignment between the two mammograms has been achieved (Maes et al., 1997). We also calculate the error against manually-defined landmarks, between each registered mammogram, and the reference mammogram. The set of landmarks are extracted from anatomical regions (nipple, visible masses and linear structures).

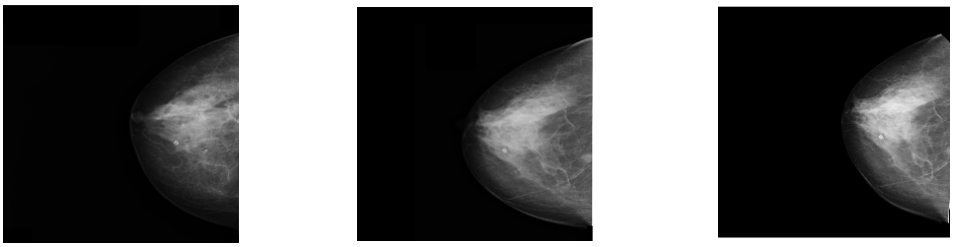
9.4.2 Mammogram registration results

We define the *oversampling factor* (OSF), which sets the number of the points of s- and t- coordinates as an integer factor of the number of the pixels along x- and y-coordinates, respectively. In our experiments, the size of the mammograms is 1024×1024 pixels (the number of the pixels along x- and y-coordinates is 1024); therefore, the number of the points of s- and t- coordinates is $1024 \times OSF$. Table 9.1 shows that the change in OSF has little effect on the performance of the proposed method. In this study, OSF is set to 2. Fig. 9.8 shows an example of a pair

Table 9.1: Effect of the OSF on registration accuracy

OSF	**1**	**2**	**3**
SSIM	0.8706	0.8773	0.871
MI	1.1753	1.1777	1.1779

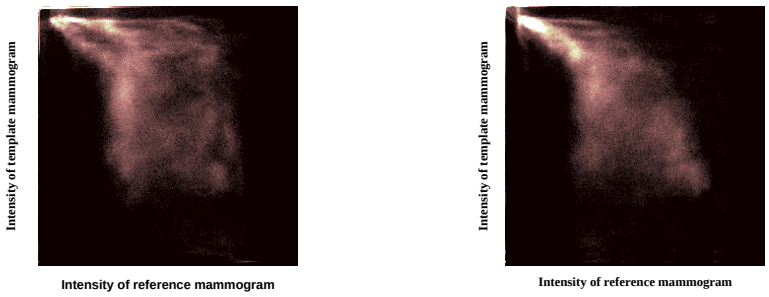
of temporal mammograms. The proposed method produced a good alignment between the reference and current (template) mammograms by compensating for the deformations between them. In addition, the proposed method aligned the local deformation without producing unrealistic results. Fig 9.9 shows the joint entropy image between the reference and current mammogram before and after registration. The entropy image after registration has less dispersion, indicating that alignment is better when the proposed method is used.



(a) **(b)** **(c)**

Figure 9.8: An example of mammogram image registration using the proposed approach: (a) the reference mammogram, (b) the current mammogram and (c) the registered mammogram

The *optimization-based* scheme is compared with a *nipple-based* scheme where the location of the nipple is manually detected so that it can be used as a reference point. As shown in Table 9.2, the mammogram registration using optimized reference points



(a) **(b)**

Figure 9.9: The joint entropy image between the reference and current mammograms, (a) before the registration, (b) after the registration using the proposed approach

improves the results of registration using nipples because it is sensitive to nipple detection accuracy and, moreover, the nipple is not visible in most mammograms. In addition, the nipple appears as a region in the mammogram, so choosing the exact reference point is problematic.

We evaluated the accuracy of the automatic scheme against the nipple-based scheme using the Wilcoxon rank sum test, and found that the SSIM and the MI values obtained with the automatic scheme vs. those obtained with the nipple-based scheme were not statistically significant (p-value$>$ 0.05).

Table 9.2: Similarity between each temporal mammogram pair before registration, using the proposed nipple-based approach and finally using the proposed approach based on an optimized reference point

Method	SSIM	MI
No registration	0.5706 ± 0.138	1.0267 ± 0.012
Nipple-based	0.7379 ± 0.011	1.1827± 0.082
Optimized reference	0.8918 ± 0.097	1.1912± 0.095

To demonstrate the effectiveness of the proposed method, we also compared it with three related methods: *Demons* (Lombaert et al., 2014), *DRAMMS* (Ou et al., 2011) and *Brandt's* method (Brandt et al., 2011).

Pectoral muscle usually appears in MLO mammograms. Several studies (e.g. (van Engeland et al., 2003)) have reported that the suppression of pectoral muscle improves the registration results. To study the effect of the pectoral muscle, we compared the registration results of the proposed approach, Demons and DRAMMS

Table 9.3: Similarity between each temporal mammogram pair with and without pectoral muscle

Method	With pectoral muscle		Without pectoral muscle	
	SSIM	*MI*	*SSIM*	*MI*
Proposed	0.8362 ± 0.0955	1.1755 ± 0.036	0.8619 ± 0.1145	1.1825 ±0.137
Demons	0.7156±0.0903	1.0630±0.0826	0.7215±0.0975	1.0471±0.0688
DRAMMS	0.8267±0.0572	1.2242±0.0981	0.8229±0.0560	1.1850±0.1170

with a set of 50 MLO mammogram pairs with and without pectoral muscle. We did not consider Brandt's method in this experiment because it can only be applied in the absence of the pectoral muscle. Table 9.3 shows the SSIM and MI values of the proposed method (with and without the pectoral muscle). The results indicate that pectoral muscle suppression improves the performance of the proposed approach. The SSIM values of Demons without the pectoral muscle were better than those with the pectoral muscle; in turn, the SSIM values of DRAMMS with and without the pectoral muscle were comparable. The MI values of the Demons and DRAMMS with the pectoral muscle were better than those without the pectoral muscle. Both SSIM and MI values without the pectoral muscle were significantly better than the values of the proposed approach with pectoral muscle (Wilcoxon rank sum test, p-value< 0.05). In the case of Demons and DRAMMS, the SSIM and MI values with and without the pectoral muscle were not statistically significant (Wilcoxon rank sum test, p-value> 0.05). With and without pectoral muscle, the SSIM values of the proposed method were of greater statistical significance than the Demons values (p-value< 0.01) and the MI values were of greater statistical significance than the DRAMMS values (p-value< 0.01).

9.4.3 Effect of boundary segmentation

Here we discuss the effect of the accuracy of boundary segmentation on the proposed method. Segmenting the breast tissue is not such a complex process because we simply separate the breast region from a black background. This process does not produce noticeable errors. Fig. 9.10 shows five segmentation scenarios of the boundaries of the reference and the template mammograms (each boundary has a different color). For each segmentation scenario, we applied the curvilinear

transformation and the registered mammogram was determined. Finally, the SSIM index between the reference and registered mammograms was calculated. Each segmentation scenario produced a different SSIM value (Fig. 9.10 (c)). The best segmentation scenario is $seg1$ (the red boundary) whereas the worst one is $seg5$ (the magenta boundary). Although $seg1$, $seg2$ and $seg3$ are different, they yielded comparable SSIM values. Indeed, the accuracy of the boundary segmentation affects the accuracy of the registration process. However, if the segmentation is good, the proposed method will provide good results. In turn, high segmentation errors may lead to very bad registration results. In the preprocessing step, we adjusted the settings of the segmentation step in order to assume a good boundary segmentation. The segmentation step did not produce such bad segmentation results as in the $seg5$ scenario.

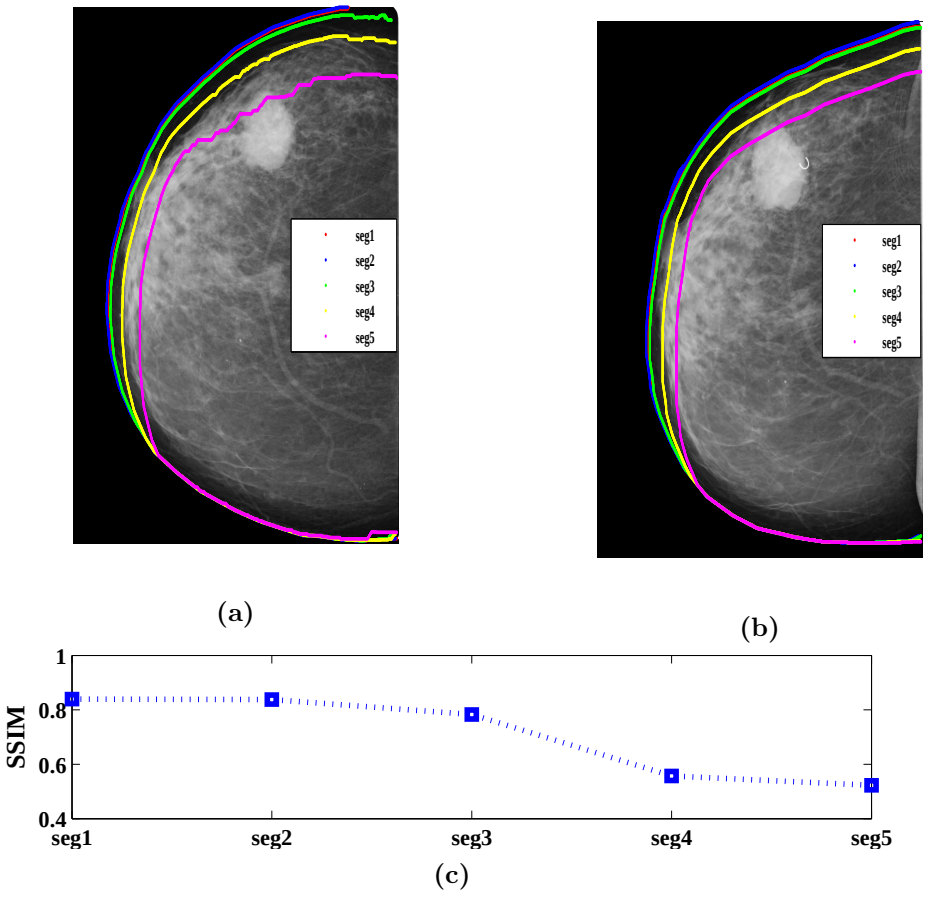


Figure 9.10: Effect of boundary segmentation, (a) the reference mammogram, (b) the template mammogram and (c) the SSIM between the reference and the registered images in each segmentation scenario

9.4.4 Comparison with other existing methods

Both Demons and DRAMMS are general-purpose medical image-registration methods while the proposed method and Brandt's method are designed to register only mammograms. *DRAMMS* is an attribute-based deformable image registration method. It describes each pixel by high-dimensional Gabor attributes extracted from multiscale and multi-orientation neighborhoods of each pixel. *Demons* is an intensity-based deformable registration approach that considers image registration as a non-parametric diffusion process. In this approach, the pixels are matched to their correspondences by local intensity characterizations.

We used the ITK implementation of Demons. Our method, as well as DRAMMS and Brandt's were implemented using the MATLAB language. In the case of DRAMMS, free form deformation was used as the deformation model and discrete optimization as the optimization strategy. In the discrete optimization, the regularization parameter was set to 0.1. In the attribute extraction, Gabor filters were used, the number of scales was set to 4 and the number of orientations was set to 6. The distance between the control points was set to 7 pixels in xy directions.

With Demons, we used the *DeformableRegistration2* ITK algorithm. The number of iterations was set to 50. The standard deviation of the Gaussian kernel which smooths the deformation field after each iteration, was set to 1.0. The aforementioned parameters produce reasonable results.

Table 9.4 presents the SSIM and MI values of the registration of 50 mammogram pairs. It also shows the error between each registered mammogram and the reference mammogram against expert-defined landmarks. The landmarks were selected from anatomical regions in the breast (nipple, visible masses and linear structures). The Euclidean distance is used to calculate the landmark errors. The proposed method yields the best SSIM values and the smallest landmark errors. The DRAMMS method gives MI values that are similar to those of the proposed method.

DRAMMS and Demons ignore the anatomical structure of the breast, as they are general purpose registration methods, whereas our method depends on the structure of the breast. The main disadvantage of Brandt's method is that it has many manual

inputs (e.g. the location of the nipple) which makes it very sensitive to the accuracy of the user. We called the time that is required to make the manual inputs MT. In contrast, the proposed method is fully automatic.

Table 9.4: Comparison between the performance of Demons, DRAMMS, Brandt's, and the proposed method

Method	SSIM	MI	Landmark errors (mm)	Time (min)
Demons	0.742 ± 0.017	1.011 ± 0.031	6.10 ± 1.32	3.7
DRAMMS	0.880 ± 0.350	1.228 ± 0.201	7.05 ± 3.32	35.3
Brandt's	0.798 ± 0.251	1.015 ± 0.092	6.93 ± 2.52	MT + 0.04
Proposed	0.903 ± 0.142	1.232 ± 0.108	5.23 ± 2.11	1.2

The results of the Wilcoxon rank sum test indicated that the landmark errors obtained with the proposed method were not significantly better than those obtained with Demons and Brandt's methods (p-value$>$ 0.05). The MI values obtained with the proposed method were of statistically greater significance than those obtained with the Demons and Brandt methods. In turn, the MI values obtained with the proposed method were not significantly better than those obtained with DRAMMS. Although the results of the proposed method and DRAMMS are comparable, the execution time of the proposed method is much shorter than for DRAMMS.

Unlike the related methods, the proposed method does not need to find control points or a correspondence algorithm. It can also cope with both global and local deformations in the breast area without causing unrealistic deformations. Table 9.5 provides a final overall comparison between the merits of the proposed method and some of the related methods. Table 9.5 shows that our method is a novel contribution, as it is fully automatic and it does not require a correspondence step. Both Demons and DRAMMS include an attribute matching step in the registration framework. In this step, an automatic approach is used to find some corresponding pixels in the reference and template images by comparing their attributes (features). The main disadvantage of Demons and DRAMMS is that the compression applied during the screening deforms the breast tissues and yields different features for the corresponding regions, which has a detrimental effect on their performance. In our experiments, the average execution time for one curvilinear transformation was 2.8 seconds. To find the optimal reference points, the exhaustive search tried 10 points

Table 9.5: Comparison to other studies

Feature	Proposed	Brandt	Demons	DRAMMS
Full Automatic	yes	no	yes	yes
Anatomical-based	yes	yes	no	no
It does not require a matching step	yes	yes	no	no
It works with CC-CC and MLO-MLO pairs	yes	no	yes	yes
It requires a segmentation step	yes	yes	no	no

(10 transformations) in the range of [0.3,0.7] in steps of 0.04. In this study, we have evaluated the performance of the proposed method using temporal mammograms from the same mammographic view (CC-CC or MLO-MLO mammographic images). The proposed method can not register mammograms that were acquired from different views (e.g. CC with MLO) because the breast has a different shape in each view. Moreover, the current version of the proposed method cannot register breast images that were acquired from different modalities (e.g. registering mammogram with MRI images for the same breast).

9.5 Conclusion

In this chapter, a new registration approach for aligning temporal mammogram images has been presented. The proposed approach uses a curvilinear coordinate system to align the mammograms. The system uses anatomical-driven coordinates instead of Cartesian coordinates, which ignore the anatomical structure of the breast. Unlike Brandt's method, which requires many manual inputs and which was designed to register only MLO-MLO mammographic pairs, the proposed method is fully automatic and it can also be applied to both CC-CC and MLO-MLO mammographic pairs. It has been shown that the representation of the curvilinear coordinates depends on its reference point. An exhaustive search is made to find the best reference point of the reference and the template mammograms which maximizes the MI between them.

The proposed registration approach was evaluated using a set of temporal mammograms. The performance of the proposed approach was compared with another scheme based on manually detected nipples. The SSIM and MI values in the case of optimized reference points were higher than these obtained by the nipple-based scheme. The SSIM and MI values of the proposed method were better than those of the three image registration methods (Demons, DRAMMS and Brandts method). It also gave the smallest landmark errors (5.23mm).

In the next chapter we propose a method for quantifying and visualizing the changes in breast tumors of patients undergoing medical treatment through strain tensors.

CHAPTER 10

Analyzing the evolution of breast tumors using flow fields and strain tensors

10.1 Introduction

Breast cancer is a group of cancer cells that can grow larger in breast tissues and may move to other areas of the human body. Mammographies are an effective method for detecting breast cancer in its early stage. Physicians usually monitor the breast tumor changes of their patients during the course of chemotherapy, and they attempt to predict pathological response in order to adjust the treatment to produce the intended effects. CAD systems help physicians to diagnose and follow-up their cases. In this chapter, we propose a computer method for quantifying and visualizing breast tumor changes in follow-up mammograms. Optical flow models with strain

tensors are proposed to make this analysis.

In the literature, several studies have been carried out to detect the changes between two successive medical images of the same subject (breast, lung, heart, etc.). Most are based on computing optical flow. Given two successive images for the same view, an optical flow algorithm computes the flow fields (i.e., displacement vectors), which map all pixels from the first image onto their new positions in the second image.

Indeed, several optical flow methods have been used in medical image analysis. Bhat and Liebling (2009) used the Lucas-Kanade (LK) method to separate the bright-field microscopy image sequence of the beating embryonic heart into two image sequences, which were then analyzed to characterize the motion of the blood and heart-wall separately. In (Wu et al., 2015; Ou et al., 2015), six deformable registration algorithms were compared to quantify tumor changes during neoadjuvant chemotherapy. The authors used Jacobian determinants and intensity residual color map to visualize the changes. Teo and Pistorius (2014) investigated the feasibility and accuracy of tracking the motion of an intruding organ-at-risk at the edges of a treatment field using a local optical flow analysis of electronic portal images. Antink et al. (2013) used a modified version of the LK method to estimate the flow fields between thoracic 4D computed tomography sequences. Krueger et al. (1998) used an optical flow method to estimate the displacement of pre- and post-compression ultrasound images. Lee (1997) used optical flow methods to align mammograms and then they compared and evaluated them to detect abnormalities.

In general, the main limitation of using optical flow methods with medical images is the lack of a ground truth for assessing their accuracy. To solve this problem, we propose using a set of robust optical flow methods with mammograms and then aggregate the best ones. This yields more confident results, since we combine the merits of these optical flow methods. In this study, we use ordered weighted averaging aggregation (OWA) operators to aggregate the results of the optical flow methods. The inputs of the proposed system are two mammograms for the same breast, one before the treatment (baseline mammogram) and one after the treatment (follow-up mammogram). To estimate the strain tensors, the system

applies five successive stages to the input mammograms (see Fig. 10.1). Firstly, five preprocessing operations are applied. Secondly, the optical flow between mammograms is computed. In this study, we assess *eight* robust optical flow methods. Thirdly, an OWA aggregation approach (Yager, 1988) is used to aggregate the results of the best optical flow methods. The aggregated flow fields are then used to calculate the strain tensors. Finally, strain tensors are shown to physicians to examine tumor changes. A *negative strain* denotes the decrease in distance between two reference points, and thus indicates a regional shrinkage. In contrast, a *positive strain* refers to a regional expansion. The proposed system produces color codes that help to visualize breast tumor changes.

10.2 Methods

Fig. 10.1 presents the five stages of the proposed system: pre-processing, computation of optical flow, aggregation, calculation of strain tensors and visualization. The subsections below give a detailed explanation of the first four steps of the proposed system.

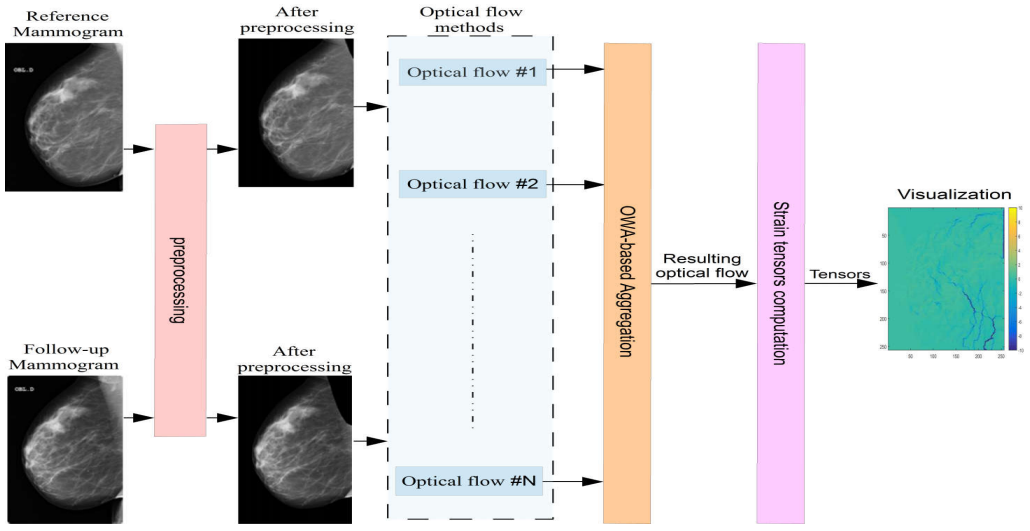


Figure 10.1: Proposed system

10.2.1 Preprocessing

The preprocessing stage includes five steps: mammogram resizing, enhancement, breast region segmentation, determination of the orientation and suppression of the pectoral muscle. We explained the steps of the preprocessing stage in section 9.3.1. The mammograms used in this study are of different sizes. In order to reduce the computation time of the optical flow methods, all the mammograms were re-sized to a resolution of 256×256 pixels. We used the detail preserving median filter proposed in (Sun and Neuvo, 1994) for noise suppression and image enhancement, and preserved the discontinuity between different regions in the breast. With this filter, the median value is computed over a square area of 5×5 pixels centered at the pixel location.

10.2.2 Calculation of the flow fields

To quantify the evolution of tumors and to reveal heterogeneous changes in disparate sub-regions within tumors, we assess eight robust optical flow methods. In the next subsections, we introduce the calculation of optical flow and briefly explain some robust optical flow methods.

10.2.2.1 LK method

The LK method assumes that the displacement of the image contents between two consecutive images is small and approximately constant within the neighborhood of each pixel (Bouguet, 2001). LK uses small neighborhoods and the least squares method to calculate the optical flow.

$$\begin{bmatrix} \sum_{i=1}^{N_p} I_{x_i}^2 & \sum_{i=1}^{N_p} I_{x_i} I_{y_i} \\ \sum_{i=1}^{N_p} I_{x_i} I_{y_i} & \sum_{i=1}^{N_p} I_{y_i}^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -\sum_{i=1}^{N_p} I_{x_i} I_{t_i} \\ -\sum_{i=1}^{N_p} I_{y_i} I_{t_i} \end{bmatrix} \tag{10.1}$$

In this equation, N_p is the number of pixels in each neighborhood, and I_x, I_y and I_t are the gradients in x-, y- and t- directions, respectively. An iterative approach

is used to solve this equation. In this study, we implemented this method to detect the large displacements that exist in mammograms.

10.2.2.2 Horn−Schunck method

The Horn−Schunck (HS) method (Horn and Schunck, 1981) is the baseline for a large family of global optical flow methods. The HS method formulates the flow as a global energy functional, and then it attempts to find the global minimum of the following objective function in terms of the displacement vector w:

$$E(w) = E(w)_{intensity} + E(w)_{smooth} \tag{10.2}$$

$E(w)_{intensity}$ is an energy term that penalizes the deviation from the brightness constancy.

$$E(w)_{intensity} = \int_{\Omega} \Psi(|I_2(x + w(x)) - I_1(x)|^2)dx \tag{10.3}$$

where $\Psi(z^2) = \sqrt{z^2 + \epsilon^2}$ is used to deal with outliers ($\epsilon = 0.001$). $E(w)_{smooth}$ is a regularity constraint used to enforce a smooth optical flow field by penalizing the total variation (TV) in the optical flow as follows:

$$E(w)_{smooth} = \int_{\Omega} \Psi(|\nabla u(x)|^2 - |\nabla v(x)|^2)dx \tag{10.4}$$

10.2.2.3 Large displacement optical flow (LDOF)

LDOF is an extension of the HS method (Brox and Malik, 2011). It uses the CTF technique and it attempts to find the global minimum of the following objective function in terms of the displacement vector w:

$$\begin{aligned} E(w) = E(w)_{intensity} + \gamma E(w)_{gradient} + \alpha E(w)_{smooth} + \\ \beta E(w, w1)_{match} + E(w1)_{descriptor} \end{aligned} \tag{10.5}$$

In this equation γ, α and β are tuning parameters. $E(w1)_{descriptor}$ is an energy term that matches a set of dense descriptors (histogram of oriented gradients; (Dalal and

Triggs, 2005)) extracted from the two mammograms. $E(w, w1)_{match}$ integrates the descriptor correspondences from a descriptor matching step (discrete process) into the variational approach (continuous process), where $w1$ is an auxiliary variable.
Due to the illumination change, the pixel constancy assumption is usually violated, so the $E(w)_{intensity}$ term does not work properly. For this reason, the $E(w)_{gradient}$ is used to add the gradient constancy constraint:

$$E(w)_{gradient} = \int_{\Omega} \Psi(|\nabla I_2(x + w(x)) - \nabla I_1(x)|^2)dx \tag{10.6}$$

where $\nabla(z)$ calculates the gradients of z in x- and y- directions of the image domain.

10.2.2.4 HOG- and MLDP- based optical flow

To cope with illumination changes, the L1-TV optical flow model was used to calculate the optical flow at point $p = (x, y)$ with the expression

$$min \underset{u,v}{E}(u, v) = \int_{\Omega} (\lambda\rho(x, y, v, u)^2 + ||\nabla u|| + ||\nabla v||) \tag{10.7}$$

where ρ is a function that measures the similarity between the images and λ is a weight. The similarity function ρ can be defined as

$$\rho(x, y, u, v) = S_2(x + u, y + v) - S_1(x, y) \tag{10.8}$$

where S_1 and S_2 are two descriptors based on HOG or MLDP extracted from the two images I1 and I2, respectively.
MLDP-based optical flow. In (Mohamed et al., 2014) the similarity function used in the data term was obtained by extracting texture features through the modified local directional pattern descriptor from two consecutive frames within the dual TV optical flow algorithm. Mohamed et al. (2014) proposed a modified version of the local directional pattern descriptor (MLDP) that assigns an 8-bit binary code to each pixel. To construct the descriptor, edge responses were calculated using the eight Kirsch compass masks for each image. For each pixel, if the edge response is

positive, the corresponding location in the binary code is set to '1'; otherwise, it is set to '0'. In this case, each mammogram is represented by eight binary channels.

HOG-based optical flow. In (Rashwan et al., 2013) the histogram of oriented gradients (HOG) is used to extract texture features from two consecutive images. HOG produces distinctive features in cases of illumination change and cluttered background (Dalal and Triggs, 2005). To construct the HOG descriptor, the occurrences of edge orientations in a local image window are counted. The image is divided into blocks (small groups of cells) and a weighted histogram is computed for each of them. The frequencies in the histograms are normalized in the interval [0,1] by using the L1-norm to compensate for illumination changes. The combination of the histograms of all blocks represents the final HOG descriptor. In the HOG-based optical flow each mammogram is represented by n channels, where each channel corresponds to one bin of the resulting histogram.

10.2.2.5 Census-based optical flow

Müller et al. (2011) proposed a global optical flow method that uses the standard total variation and census as a texture descriptor in the data term. The census operator labels the pixels of an image by comparing the elements of the $k_c \times k_c$ neighborhood surrounding each pixel with the value of this pixel (Ojala et al., 2002). Pixels in this window with a value greater than the central pixel are labeled as '1' and the rest as '0'; thus, if k_c is set to 3, each pixel is represented by 8 bits.

10.2.2.6 Classic+NL optical flow

The *Classic+NL* method (Sun et al., 2010) is a modified and optimized version of the HS method. To determine the optical flow it uses a median filter-based non-local

term, in addition to the CTF technique and it optimizes the following energy:

$$\begin{aligned} E_O(u,v) = & \sum_{i,j}\{\Psi_d(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \\ & +\lambda_1[\Psi_s(u_{i,j} - u_{i+1,j}) + \Psi_s(u_{i,j} - u_{i,j+1}) \\ & +\Psi_s(v_{i,j} - v_{i+1,j}) + \Psi_s(v_{i,j} - v_{i,j+1})] \\ & +\lambda_2(\|u-\hat{u}\|^2 + \|v-\hat{v}\|^2)\} \end{aligned} \tag{10.9}$$

$$\begin{aligned} E_C(\hat{u},\hat{v}) = & \lambda_2(\|u-\hat{u}\|^2 + \|v-\hat{v}\|^2) + \\ & \underbrace{\sum_{i,j}\sum_{(q,z)\in N_{i,j}} \omega_{i,j,q,z}\left(|\hat{u}_{i,j} - \hat{u}_{q,z}| + |\hat{v}_{i,j} - \hat{v}_{q,z}|\right)}_{\text{non-local term}} \end{aligned}$$

$E_O(u,v)$ is the main energy function and $E_C(\hat{u},\hat{v})$ is the proposed non-local term used to refine the resulting optical flow of Eq. 10.9. $\hat{u}$ and $\hat{v}$ are auxiliary flow fields. $N_{i,j}$ is a set of neighbors of pixel (i,j), and λ_1 and λ_2 are scalar weights. For each pixel, a non-local term is used to determine which pixels in the area belong to the same surface. Subsequently, they are weighted using $\omega_{i,j,q,z}$ which determines the probability of pixel (q,z) existing in the same surface of pixel (i,j). In this way, the details of the optical flow at the boundaries of the regions are preserved. In the case of *Classic+NL*, all values of $\omega_{i,j,q,z}$ are equal and belong to the edges of the regions. Below we refer to *Classic+NL* by CL-NL and *Classic+NL-FULL* by CL-Full.

10.2.3 Aggregating optical flow models using OWA operators

As we mentioned above, the main problem in evaluating the performance of optical flow methods is the lack of ground truth. A landmark-based error can give us a provisional evaluation for the optical flow methods. To estimate the optical flow accurately, we selected some recent optical flow methods. Indeed, each optical flow method has its own characteristics. For example, the HOG, MLDP and LDOF methods are based on feature extraction and they are more robust against

illumination changes. In turn, the standard methods, such as HS or CL-NL, are based on intensity or color values that hold all information about image pixels and image sensors. In this study we aggregated the displacements obtained using OWA operators. We assessed three OWA operators: *'at least half'*, *'most of them'* and *'as many as possible'*. We used the optical flow obtained from the best aggregation method to calculate the strain tensors. In the following subsections, we introduce the OWA operators and the method used to aggregate the results of the optical flow methods.

10.2.3.1 OWA operators

Yager (1988) used the OWA operators to perform several aggregation tasks. Given an input vector $f = [f_1, f_2, \ldots, f_n]$ and a weighting vector $w = [w_1, w_2, \ldots, w_n]$, the aggregated value of f can be calculated as follows:

$$F(f_1, f_2, \ldots, f_n) = \sum_{k=1}^{n} w_k r_k \tag{10.10}$$

where $\sum_{k=1}^{n} w_k = 1$, n is the number of elements of f and r_k is the kth largest element of f, $k = 1, 2, \ldots, n$.

To calculate the weights we used the increasing linguistic quantifier Q proposed by Yager (1993). They can be calculated as follows:

$$Q(r) = \begin{cases} 0 & \text{if } r < 0 \\ \frac{(r-a)}{b-a} & \text{if } a \leq r \leq b \\ 1 & \text{if } r > b \end{cases} \tag{10.11}$$

Using the values of a and b, we can define different increasing quantifiers. To calculate 'at least half', 'most of them' and 'as many as possible', (a, b) are set to $(0, 0.5)$, $(0.3, 0.8)$ and $(0.5, 1)$, respectively. For a specific quantifier, the weights can be

calculated as follows:

$$w_k = Q\left(\frac{k}{n}\right) - Q\left(\frac{k-1}{n}\right) \quad for\ k = 1, \ldots, n \tag{10.12}$$

Fig. 10.2 shows these three quantifiers.

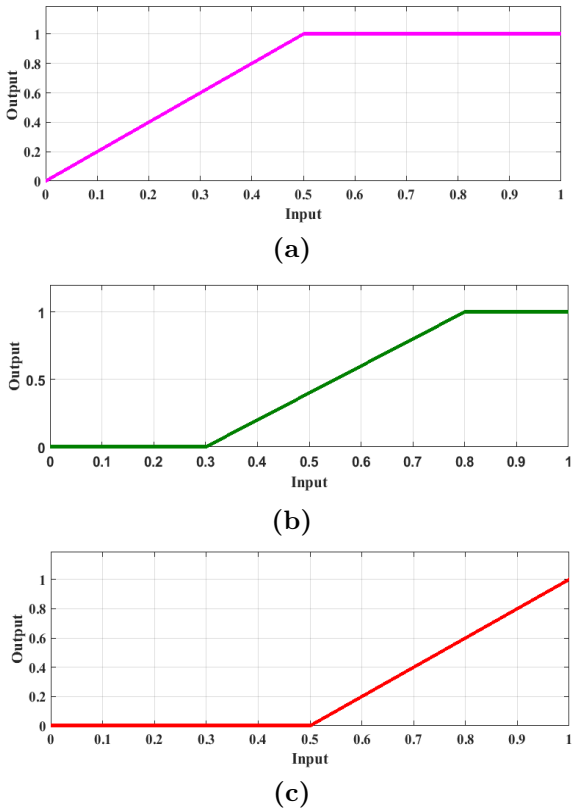


Figure 10.2: The linguistic quantifiers (a) 'at least half', (b) 'most of them' and (c) 'as many as possible'

10.2.3.2 Aggregating optical flow models

Given L optical flow methods, for each pixel in the image (i, j), the qth optical flow method produces the displacement field $(u_q(i, j), v_q(i, j))$. We apply the aggregation function on the values of u and v separately. For instance, if there are three optical flow methods ($L = 3$), for each pixel, the optical flow methods produce three displacement vectors (u_1, v_1), (u_2, v_2) and (u_3, v_3). The aggregation operator aggregates u_1, u_2 and u_3 into one value u_{agg}. In the same way, v_1, v_2 and v_3 are aggregated into one value v_{agg}. Therefore, for each pixel in the image, the aggregation

operator produces the aggregated displacement field (u_{agg},v_{agg}). Thus, aggregating optical flow methods using 'at least half', 'most of them' and 'as many as possible' produces (U_{least}, V_{least}), (U_{most}, V_{most}) and (U_{many}, V_{many}), respectively.

10.2.4 Calculation of the strain tensors

Strain is the change in length per length unit. A negative strain denotes a shrinkage while a positive strain denotes an expansion. It can be defined using the displacement gradient with the expression (Abd-Elmoniem et al., 2008)

$$\nabla w(p,t) = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \tag{10.13}$$

where $p = [x, y]^T$ are the coordinates of each pixel in the image and u and v are the aggregated flow fields. Then the deformation gradient F is given by

$$F(p,t) = (\mathbf{I} - \nabla w(p,t))^{-1} \tag{10.14}$$

where $\mathbf{I}$ is the identity matrix. Finally, the strain tensor is defined as

$$s(p,t) = \frac{1}{2}\left[\mathbf{I} - (F^{-1}(p,t))^T F^{-1}(p,t)\right] \tag{10.15}$$

$$s(p,t) = \begin{bmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{bmatrix} \tag{10.16}$$

where s_{xx}, s_{xy}, s_{yx} and s_{yy} are the components of the strain tensor. s_{xx} is the unit elongation for an element originally in the x-direction. s_{yy} is the unit elongation for an element originally in the y-direction. s_{xy} gives the decrease in angle between two elements initially in the x- and y- directions. Note that the components of the strain tensor are the same size as the mammograms. s_{xy} and s_{yx} are identical, so we only show s_{xy} in our experiments.

10.3 Experimental results and discussion

A dataset of 34 mammogram pairs was used in our experiments. It was collected from the Hospital Universitari Sant Joan (Reus, Spain). The temporal mammograms are *Full Field Digital Mammograms (FFDM)* with sizes of 1770×2370, 3540×4740 and 2364×2964 pixels, and with a spatial resolution of 100μ.

To evaluate the performance of the optical flow methods, we calculate the error between each registered follow-up mammogram and the baseline mammogram against manually-defined landmarks. The extracted landmarks are in corresponding anatomical regions (nipple, visible masses and linear structures). The number of landmarks varies from one mammogram pair to another. Assume that L_{base} is the list that contains the positions of the landmarks picked from the baseline mammogram, and L_{fol} contains the corresponding positions of the landmarks in the follow-up mammogram. The Euclidean distance is used to calculate the landmark error.

Statistical analysis may provide more evidence when analyzing optical flow methods, so we calculate the statistical significance between the landmark error of the optical flow methods using the *Wilcoxon signed rank test*.

10.3.1 Analysis of optical flow methods

In this subsection we analyze the performance of eight optical flow methods. We experimentally tuned the parameters of each optical flow method, which are summarized in the following paragraph.

Parameter configuration. The parameters of *LK* were 9 pyramid levels, a 9×9 neighborhood and 30 iterations. In the case of *HS*, image pyramids were constructed using the configuration proposed in (Bruhn et al., 2005), there were 10 warpings per image pyramid level and 5 pyramids. With *LDOF*, each HOG descriptor was calculated in 15 different orientations and computed in 7×7 neighborhoods. The final HOG descriptors were calculated by combining the histograms at the central pixel and the eight neighbors at a distance of four pixels. Parameters γ, α and β were 5, 30 and 300, respectively. There were 92 levels in the image pyramid. This number

was selected so that the discrete derivative filters could still be applied at the top of the pyramid. In the case of *CL-NL and CL-Full*, the number of levels of image pyramids in the first stage of graduated non-convexity (GNC) optimization was 4, while it was 2 in the second stage of GNC. There were 10 warpings per pyramid level. The half window size of the weighted median filter was 10 and the size of the median filter was 5×5 pixels. With *OF-Census*, the census descriptor was calculated in a 3×3 neighborhood around each pixel. With *OF-HOG*, the HOG was computed in a 5×5 neighborhood with 8 orientations. With *OF-MLDP*, the MLDP were calculated using eight Kirsch compass masks convoluted with a Gaussian kernel to reduce the effect of illumination change. With OF-Census, OF-HOG and OF-MLDP, there were 10 warps per level and 64 levels in the image pyramid. The weighting of the data term was 10.

The experiments were carried out using MATLAB on an Intel processor core 2 Quad at 2.5GHz and 8 GB of RAM. In this study, we use state-of-the-art optical flow methods. Their implementations are freely available. The implementations of CL-NL, CL-Full and HS are available at http://cs.brown.edu/people/black/code.html/. The implementations of OF-HOG, OF-MLDP and OF-Census can be requested from the authors of (Rashwan et al., 2013). The implementation of LDOF is available at http://www.cs.berkeley.edu/~katef/LDOF.html/.

Evaluation. Table 10.1 presents the evaluation of the optical flow methods using

Table 10.1: Evaluating optical flow methods

Method	Landmarks error	p-value	Significance
LK	4.92±1.61	0.0011	yes
HS	4.58±2.65	0.0573	no
LDOF	6.64±4.40	0.00091568	yes
OF-HOG	4.27±2.00	0.0553	no
OF-MLDP	6.23±4.24	0.00092726	yes
OF-Census	6.05±3.30	0.0014	yes
CL-NL	**3.33±1.81**	**-**	**-**
CL-Full	4.69±2.56	0.0013	yes

the mean and the standard deviation of the landmark error. It also shows the results of the Wilcoxon signed rank test. CL-NL gives the smallest landmark error (the best

performance) followed by OF-HOG and HS. LDOF has the biggest landmark error (both mean and standard deviation of landmark error) and, therefore, performs worst. The results in Table 10.1 enable the optical flow methods studied to be grouped in three categories as follows. The first category contains only the CL-NL method, which has a landmark error smaller than 4. The second category contains the LK, HS, OF-HOG and CL-Full methods, which show a landmark error smaller than 5. The third category contains the LDOF, OF-Census and OF-MLDP methods, which have a landmark error greater than 6. Obviously, CL-NL is the best optical flow method because it produces the smallest landmark error. Cl-NL uses a weighted median filter (in the non-local term) to remove the outliers from the calculated optical flow in each level of the CTF, and this makes the resulting optical flow more accurate. As CL-NL is the optical flow method that performs best we measure the statistical significance of its results using other methods and the Wilcoxon signed rank test. In the third column of Table 10.1, a p-value lower than 0.05 means that there is a significant difference with respect to the CL-NL method. In the fourth column, 'yes' means that the errors have a significant difference. As shown in Table 10.1, the results of CL-NL are significantly better than the results of LK, LDOF, OF-MLDP, OF-Census and CL-Full (p-value < 0.05). On the contrary, the results of CL-NL are not different enough from the results of the HS and OF-HOG methods. This means that the performances of HS and OF-HOG are similar to the performance of CL-NL. These findings prompt us to aggregate CL-NL, HS and OF-HOG (the top three methods) in the next stage of the proposed system.

Fig. 10.3 presents the average execution time of the optical flow methods. HS determines the optical flow in the shortest average execution time followed by LDOF. CL-Full takes longest to determine the optical flow. CL-NL and OF-HOG take less than 5 minutes. The best three optical flow methods (CL-NL, HS and OF-HOG) take less time than the CL-FULL and LK to determine the optical flow. Thus, the optical flow methods that are to be aggregated are also good in terms of execution time.

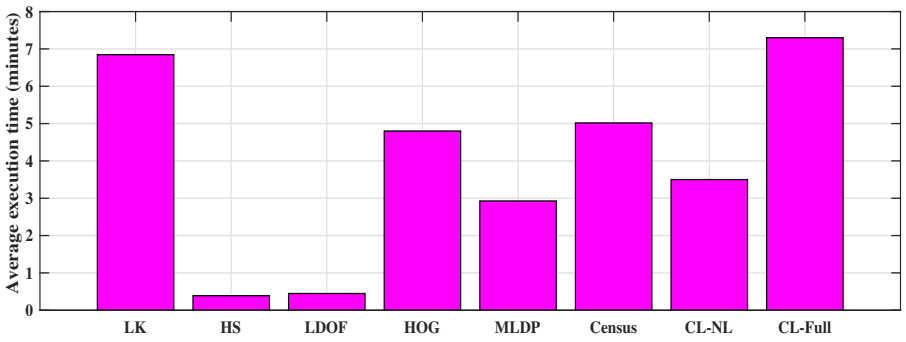


Figure 10.3: The average execution time of the optical flow methods

10.3.2 Analysis of the aggregation methods

Here we analyze the performance of three aggregation methods based on OWA operators: 'at least half', 'most of them' and 'as many as possible'. Below, we refer to these operators as 'least', 'most' and 'many', respectively. We selected the top *three* optical flow methods to be aggregated: CL-NL, OF-HOG and HS. We also used the *Wilcoxon signed rank test* and the landmark error to evaluate the results of the aggregation methods.

Table 10.2 shows the landmark error of the aggregation methods. Optical flow aggregation using the 'many' approach gives the smallest landmark error (2.92±1.38) followed by the 'least' approach. The worst aggregation results are given by the 'most' operator. The mean and the standard deviation of the landmark error are smaller for 'many' than for CL-NL. This means that 'many' more accurate than the other aggregation methods or the individual use of any other optical flow method.

As the 'many' aggregator is the best option, we measure the statistical significance of its results in comparison with the 'least' and the 'most'. As shown in Table 10.2, the results of 'many' are significantly better than the ones of 'least' and 'most'.

Table 10.2: Results of the aggregated optical flow models

Method	least	most	many
Landmark-error	4.28±2.34	5.56±3.61	**2.92±1.38**
p-value	0.01	8.21E-004	-
significant	yes	yes	-

We also calculate the statistical significance of the results of the 'least', 'most' and 'many' approaches in comparison with the results of CL-NL. As shown in Table 10.3,

the results of CL-NL are significantly better than the results of the 'most' approach (p-value < 0.05), but not better than the results of the other two aggregators (p-value > 0.05).

Table 10.3: Statistical analysis of the aggregation approaches

CL-NL vs.	least	most	many
p-value	0.08	1.56E-002	0.35
significant	no	yes	no

10.3.3 Analysis of tumor changes

On the basis of the findings of the above experiments, we use the aggregated optical flow of 'many' to calculate the strain tensors. The strain tensors at a given pixel indicate the change between the follow-up and the baseline mammograms: >0 for expansion, <0 for shrinkage, and 0 for no change. Below, we present two examples of the calculation of strain tensors using the proposed method. In both examples, the baseline mammogram was acquired before the treatment. The follow-up mammogram of the first example was acquired after a short period of treatment. The follow-up mammogram of the second example was acquired after a long period of treatment. The four components of the strain tensor have the same size of mammograms. Although the strain tensor has four components (s_{xx}, s_{xy}, s_{yx} and s_{yy}), we show only three of them in the figures because s_{xy} and s_{yx} are identical. Fig. 10.4 shows the strain tensors based on the 'many' approach. The period between the baseline and the follow-up mammograms was 49 days. In this case, the landmark error is 4.8012 pixels. As shown in Fig. 10.4, at each point, the color map of the components of the strain tensor (s_{xx}, s_{xy} and s_{yy}) provides a quantification of the tumor changes. In particular, they show expansions in the middle and shrinkage in the bottom-left side of the breast.

To summarize the color maps of the three strain tensors (s_{xx}, s_{xy} and s_{yy}) on one color map, we calculate the maximum absolute eigenvalue (λ_{max}) of the strain tensor at each pixel $s(p,t)$ in the mammogram. Fig. 10.4 (f) shows λ_{max}. Some shrinkage in the middle and in the bottom-left side of the breast clearly appear in λ_{max}. Fig. 10.5

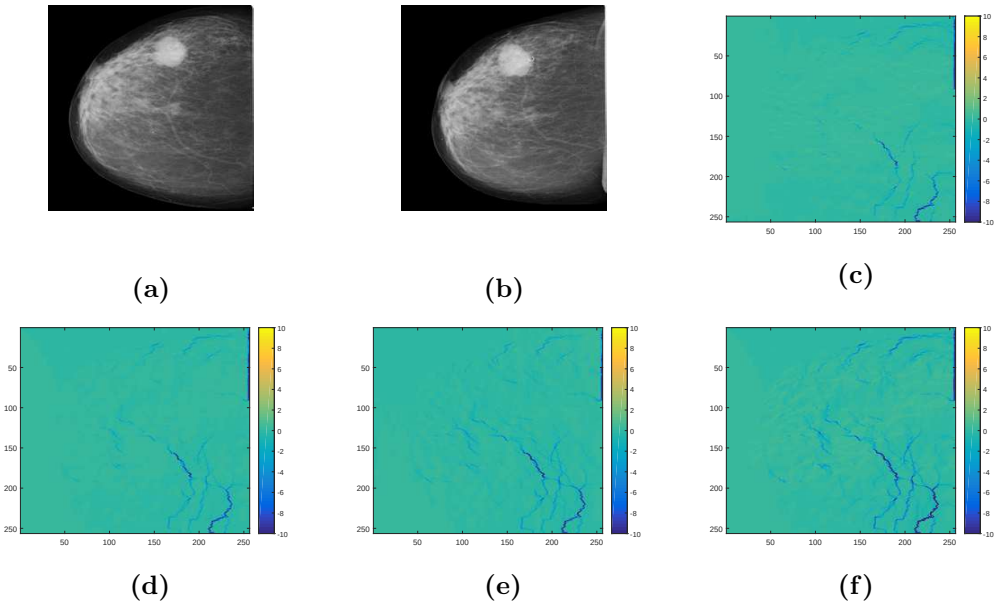


Figure 10.4: The calculated strain tensors using the 'many' approach for a patient undergoing short-term treatment: (a) the baseline mammogram, (b) the follow-up mammogram, (c) s_{xx}, (d) s_{xy}, (e) s_{yy} and (f) λ_{max}

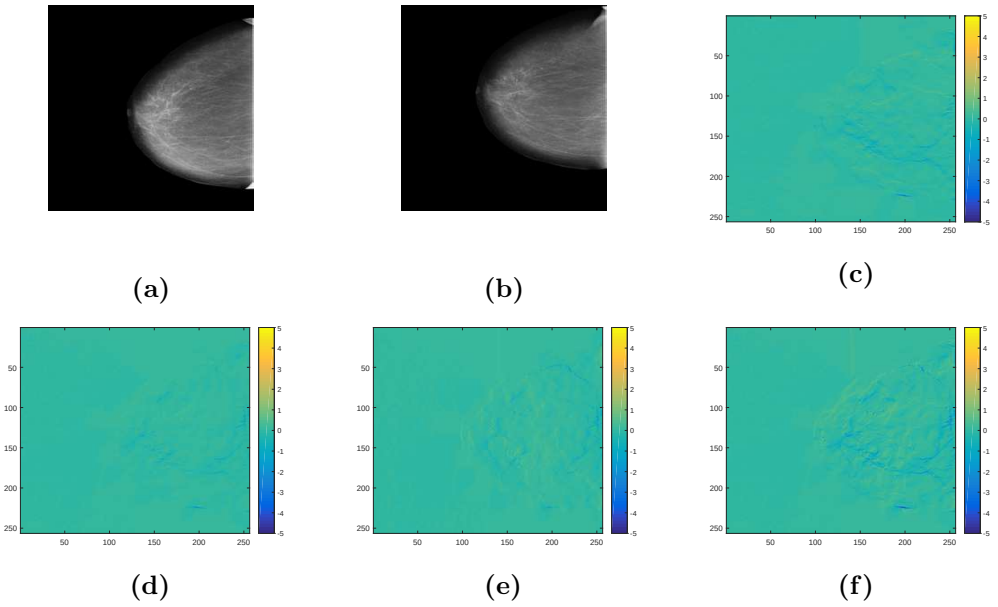


Figure 10.5: The calculated strain tensors using the 'many' approach for a patient undergoing long-term treatment: (a) the baseline mammogram, (b) the follow-up mammogram, (c) s_{xx}, (d) s_{xy}, (e) s_{yy} and (f) λ_{max}

shows the calculated strain tensors for a patient undergoing long-term treatment (415 days). In this case, the landmark error is 1.7619 pixels. As shown, the strain tensors can help to visualize breast tumor changes. The λ_{max} color map shows little

expansion in the middle of the breast of this patient.

10.4 Conclusion

In this chapter we have proposed a computer method for quantifying and visualizing the changes in breast tumors for patients undergoing medical treatments using strain tensors. The proposed method consists of five successive stages: pre-processing, calculation of the optical flow, aggregation, calculation of strain tensors and visualization. We determined the displacement fields between each follow-up mammogram and its baseline. The displacement fields obtained were then used to calculate the strain tensors. We evaluated the performance of eight optical flow methods using the landmark error and the Wilcoxon signed rank test. We found that CL-NL produces the smallest landmark error (best performance) followed by HS and OF-HOG. Cl-NL accurately estimates optical flow because it uses a weighted median filter to remove the outliers from the calculated optical flow on each level of the CTF.

As there is no ground truth for evaluating optical flow methods when they are applied to mammograms, we propose to aggregate the top three optical flow methods using OWA operators. To do so, we assessed three OWA-based aggregation approaches: 'at least half', 'most of them' and 'as many as possible'. The aggregated optical flow methods using the 'as many as possible' operator produced smaller landmark error than the method using CL-NL, which indicates that this aggregation approach improves the accuracy of the optical flow calculation. As a result, the optical flow obtained in the 'as many as possible' approach is used to calculate the strain tensors. The proposed system displays the strain tensors as well as the maximum absolute eigenvalue (λ_{max}). The proposed method provides a good quantification and visualization of breast tumor changes and it may help physicians to plan treatment for their patients.

In the next section we conclude the thesis and present some lines of future research.

Part VI

Conclusion

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT OF ADVANCED COMPUTER METHODS FOR BREAST CANCER IMAGE INTERPRETATION THROUGH TEXTURE AND TEMPORAL EVOLUTION ANALYSIS
Mohamed Abdelnasser Mohamed Mahmoud

CHAPTER 11

Concluding remarks

11.1 Summary of contributions

In this thesis, we have analyzed breast tissues in three imaging modalities: mammography, ultrasonography and thermography. Our analysis includes mass/normal breast tissue classification, benign/malignant tumor classification in mammograms and ultrasound images, nipple detection in thermograms, mammogram registration and analysis of the evolution of breast tumors. The thesis is divided into six parts: introduction (chapters 1 and 2), analysis of breast cancer in mammograms (chapters 3, 4 and 5), analysis of breast cancer in ultrasound images (chapters 6 and 7), analysis of breast cancer in infrared images (chapter 8) and analysis of the temporal evolution of breast cancer (chapters 9 and 10).

In chapter 3, we compared several texture analysis methods for breast mass detection, using images from a public breast cancer database. In particular, we proposed LDN

as a new feature extraction method in this field. LDN improved the results of such well-known texture analysis methods as LBP, HOG, GLCM or Gabor filters. To improve mass detection rates, we proposed using two combination schemes. Firstly, we concatenated the features of the best texture analysis methods. Of all the concatenations, LDN+LBP gave the best overall results. Secondly, we used the classifier voting technique to combine the predictions given by LDN, LBP, RLBP and HoG. Results were good. The $LDN + LBP$ combination detected the false positives of breast mass detection; as a result, the number of unnecessary biopsies can be reduced.

In chapter 4, we proposed a CAD system for analyzing breast tissues in mammograms. This analysis included breast tissue classification and breast density classification. We proposed the ULDP descriptor for feature extraction. On the basis of our experiments in chapter 4, we can conclude the following:

- Results were good with both mass/normal breast tissue classification and breast tissue density classification. In addition, breast tissue density classification based on the histogram of ULDP yielded a high kappa coefficient.
- The correct classification rate based on the ULDP descriptor with digital mammograms was higher than the one with film-screen mammograms, due to the good contrast of digital mammograms.
- We studied the effect of breast density on the performance of the proposed descriptor. ULDP performed more or less constantly with different breast densities.

In chapter 5, we studied the effect of factors such as pixel resolution, integration scale, preprocessing and feature normalization on the performance of these texture methods for mass classification. Texture analysis methods, when applied to *benign/malignant* mass classification in mammograms, are sensitive to these factors. The best combination of these factors should be identified to achieve the best discriminative power of each texture analysis method. We expect that the assessment performed in this study will help researchers to do so. Due to its computational cost advantage, sequential forward selection would be a suitable approach for determining

a reasonable (possibly the best) factor configuration.

In chapter 6, FLDP is proposed for breast tissue characterization in ultrasound images. FLDP describes each pixel in a given image by its edge responses and makes use of fuzzy membership functions. LSVM and NLSVM classifiers are used to demonstrate the effectiveness of FLDP at discriminating between mass and normal tissues. The results show that the proposed descriptor leads to better results than some of the state-of-the-art descriptors.

In chapter 7, we proposed the use of super-resolution approaches, which exploit the complementary information provided by multiple images of the same target. The proposed CAD system consists of four stages: super-resolution computation, extraction of the region of interest, feature extraction and classification. We evaluated the performance of five texture methods with the proposed system: gray level co-occurrence matrix features, local binary patterns, phase congruency-based local binary pattern, histogram of oriented gradients and pattern lacunarity spectrum. We showed that our super-resolution-based approach improves the performance of the evaluated texture methods and thus outperforms the state of the art in benign/malignant tumor classification.

An automatic, accurate and real-time method for detecting nipples in thermograms was proposed in chapter 8. The main stages of the proposed method are: human body segmentation, determination of nipple candidates using adaptive thresholding and detection of the nipples using the proposed nipple selection algorithm.

A registration approach for aligning temporal mammogram images was presented in chapter 9. The proposed approach uses a curvilinear coordinate system to align the mammograms. In this way the system uses anatomical-driven coordinates instead of Cartesian coordinates, which ignore the anatomical structure of the breast. The proposed method is fully automatic and it can also be applied to both CC-CC and MLO-MLO mammographic pairs. The performance of the approach was compared with another scheme based on manually detected nipples. The SSIM and MI values in the case of optimized reference points were higher than these obtained in the nipple-based scheme. The method gave better registration results than three image

registration methods (Demons, DRAMMS and Brandt's method).

In chapter 10, we proposed a computer method for quantifying and visualizing the changes in breast tumors for patients undergoing medical treatments using flow fields and strain tensors. In general, the main limitation of using optical flow methods with medical images is the lack of a ground truth to asses their accuracy. To solve this problem, we proposed using a set of robust optical flow methods with mammograms and then aggregating the best ones. This yielded more confident results, since we tend to combine the merits of these optical flow methods. We have used OWA operators to aggregate optical flow methods. The proposed method provides good quantification and visualization of the evolution of breast tumor, and it may help physicians to plan treatment for their patients.

11.2 Future research lines

The work presented in this thesis makes a contribution to the interpretation of breast cancer image analysis. We believe this is an interesting and important field of research. Several directions of future work have been identified during this work. For example:

1. Deep learning techniques for analyzing breast cancer images. First, we will explore the use of deep convolution neural network and autoencoders to analyze breast ultrasound images. Second, we will apply unsupervised feature learning techniques to analyze dynamic thermograms.
2. Combination schemes to improve the performance of a computer-aided diagnostic system based on multiple mammographic views and different breast images modalities, such as magnetic resonance and ultrasound.
3. ULDP to predict the malignancy of the suspicious ROIs in mammograms (classification of the mass ROIs into benign or malignant). We will also use ULDP to segment dense regions from the breast area, then we will calculate the breast density as a percentage of the area of the dense tissue and the area of the whole breast. Given the calculated percentage, we will determine the

corresponding BI-RADS category. Moreover, we will use the ULDP to classify the breast density of each ultrasound image according to BI-RADS categories.

4. A biopsy procedure is usually used to discriminate between benign and malignant masses. Although biopsies are expensive, they are often carried out unnecessarily. To reduce this cost, we will implement a CAD system to improve the results of *benign/malignant* mass classification. Given a mammogram and an ultrasound image for each patient, we will aggregate the ULDP features extracted from the suspicious ROIs. Then we will use these features to classify the masses into benign or malignant.
5. Extension of the FLDP descriptor by using higher order membership functions such as Gaussian and Trapezoidal functions. In addition, we will use neutrosophic logic instead of fuzzy logic. Neutrosophic logic is a general framework for unifying many existing logics including fuzzy logic. Thus, principles such as neutrosophic sets and neutrosophic probability will be used instead of fuzzy sets and the degree of membership.
6. Integration of the nipple detection method with such applications as breast region segmentation and finding similar masses in mammograms and thermograms of the same patients. In addition, we will use the proposed nipple detection method to find the optimal location of the origin of the curvilinear coordinates, which will be used to register thermograms.
7. Adaptation of the curvilinear registration approach for multi-modal mammogram image registration (for example, registering mammograms and MRI images for the same breast). In addition, we will evaluate the performance of the proposed approach with *tomosynthesis* image registration.
8. The use of the flow aggregation approach to estimate strain tensors in breast ultrasound elastography. Moreover, the optical flow aggregation approach can be used in several applications, specially those that require displacement fields to be calculated, such as medical image super-resolution and image registration.

References

Abd-Elmoniem, K. Z., Stuber, M., and Prince, J. L. (2008). Direct three-dimensional myocardial strain tensor quantification and tracking using zharp. *Medical Image Analysis*, 12(6):778–786.

Agarwal, D., Shriram, K., and Subramanian, N. (2013). Automatic view classification of echocardiograms using histogram of oriented gradients. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pages 1368–1371. IEEE.

Ahonen, T. and Pietikäinen, M. (2007). Soft histograms for local binary patterns. In *Proceedings of the Finnish signal processing symposium, FINSIG*, volume 5, page 1.

Aksoy, S. and Haralick, R. M. (2001). Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, 22(5):563–582.

Ali, M. A., Sayed, G. I., Gaber, T., Hassanien, A. E., Snasel, V., and Silva, L. F. (2015). Detection of breast abnormalities of thermograms based on a new segmentation method. In *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 255–261. IEEE.

Allain, C. and Cloitre, M. (1991). Characterizing the lacunarity of random and deterministic fractal sets. *Phys. Rev. A*, 44:3552–3558.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.

Anand, S., Kumari, R. S. S., Jeeva, S., and Thivya, T. (2013). Directionlet transform based sharpening and enhancement of mammographic X-ray images. *Biomedical Signal Processing and Control*, 8(4):391–399.

Angulo, A., Ferrer, J., Pinto, J., Lavarello, R., Guerrero, J., and Castaneda, B. (2015). Experimental assessment of an automatic breast density classification algorithm based on principal component analysis applied to histogram data. In *Tenth International Symposium on Medical Information Processing and Analysis*, pages 92870E–92870E. International Society for Optics and Photonics.

Antink, C. B. H., Singh, T., Singla, P., and Podgorsak, M. (2013). Evaluation of advanced Lukas–Kanade optical flow on thoracic 4D-CT. *Journal of Clinical Monitoring and Computing*, 27(4):433–441.

Armitage, P., Berry, G., and Matthews, J. N. (2008). *Statistical Methods in Medical Research*. John Wiley & Sons.

Ayyala, R. S., Chorlton, M., Behrman, R. H., Kornguth, P. J., and Slanetz, P. J. (2008). Digital mammographic artifacts on full-field systems: What are they and how do I fix them? 1. *Radiographics*, 28(7):1999–2008.

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, 22(4):469–483.

Bellotti1, R., Carlo, F. D., Tangaro, S., Gargano, G., Maggipinto, G., Castellano, M., Massafra, R., Cascio, D., Fauci, F., Magro, R., Raso, G., Lauria, A., Forni, G., Bagnasco, S., Cerello, P., Zanon, E., Cheran, S. C., Torres, E. L., Bottigli, U., Masala, G. L., Oliva, P., Retico, A., Fantacci, M. E., Cataldo, R., Mitri1,

I. D., and Nunzio, G. D. (2006). A completely automated CAD system for mass detection in a large mammographic database. *Medical physics*, 33(8):3066–3075.

Beura, S., Majhi, B., and Dash, R. (2015). Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing*, 154:1–14.

Bhat, S. and Liebling, M. (2009). Cardiac tissue and erythrocyte separation in bright-field microscopy images of the embryonic zebrafish heart for motion estimation. In *Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, June 28 - July 1, 2009*, pages 746–749.

Borchartt, T. B., Conci, A., Lima, R. C., Resmini, R., and Sanchez, A. (2013). Breast thermography from an image processing viewpoint: A survey. *Signal Processing*, 93(10):2785–2803.

Boucher, A., Cloppet, F., Vincent, N., and Jouve, P. (2010). Visual perception driven registration of mammograms. In *20th International Conference on Pattern Recognition (ICPR)*, pages 2374–2377. IEEE.

Bouguet, J.-Y. (2001). Pyramidal implementation of the affine Lucas Kanade feature tracker description of the algorithm. *Intel Corporation*, 5:1–10.

Brandt, S. S., Karemore, G., Karssemeijer, N., and Nielsen, M. (2011). An anatomically oriented breast coordinate system for mammogram analysis. *IEEE Transactions on Medical Imaging*, 30(10):1841–1851.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brown, L. G. (1992). A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376.

Brox, T. and Malik, J. (2011). Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513.

Bruhn, A., Weickert, J., and Schnörr, C. (2005). Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231.

Cabral, T. M. and Rangayyan, R. M. (2012). Fractal analysis of breast masses in mammograms. *Synthesis Lectures on Biomedical Engineering*, 7(2):1–118.

Cai, L., Wang, X., Wang, Y., Guo, Y., Yu, J., and Wang, Y. (2015). Robust phase-based texture descriptor for classification of breast ultrasound images. *Biomedical engineering online*, 14(1):26.

Campanini, R., Dongiovanni, D., Iampieri, E., Lanconelli, N., Masotti, M., Palermo, G., Riccardi, A., and Roffilli, M. (2004). A novel featureless approach to mass detection in digital mammograms based on support vector machines. *Physics in Medicine and Biology*, 49(6):961.

Chakraborty, J., Midya, A., Mukhopadhyay, S., Rangayyan, R. M., Sadhu, A., Singla, V., Khandelwal, N., Bhattacharyya, P., and Azevedo-Marques, P. M. (2015). Detection of the nipple in mammograms with gabor filters and the radon transform. *Biomedical Signal Processing and Control*, 15:80–89.

Chan, H.-P., Sahiner, B., Petrick, N., Helvie, M. A., Lam, K. L., Adler, D. D., and Goodsitt, M. M. (1997). Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network. *Physics in Medicine and Biology*, 42(3):549.

Chandrasekhar, R. and Attikiouzel, Y. (1996). Gross segmentation of mammograms using a polynomial model. In *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, volume 3, pages 1056–1058. IEEE.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Chen, D.-R., Chang, R.-F., Chen, C.-J., Ho, M.-F., Kuo, S.-J., Chen, S.-T., Hung, S.-J., and Moon, W. K. (2005). Classification of breast ultrasound images using fractal feature. *Clinical imaging*, 29(4):235–245.

Chen, J., Kellokumpu, V., Zhao, G., and Pietikäinen, M. (2013). Rlbp: Robust local binary pattern. In *Proc. the British Machine Vision Conference (BMVC 2013), Bristol, UK*.

Christoyianni, I., Koutras, A., Dermatas, E., and Kokkinakis, G. (2002). Computer aided diagnosis of breast cancer in digitized mammograms. *Computerized Medical Imaging and Graphics*, 26(5):309–319.

Chu, J., Min, H., Liu, L., and Lu, W. (2015). A novel computer aided breast mass detection scheme based on morphological enhancement and SLIC superpixel segmentation. *Medical Physics*, 42(7):3859–3869.

Clausi, D. A. (2002). An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62.

Crocker, J., Jones, E., and Curran, R. (1983). A comparative study of nuclear form factor, area and diameter in non-hodgkin's lymphomas and reactive lymph nodes. *Journal of Clinical Pathology*, 36(3):298–302.

Crum, W. R., Hartkens, T., and Hill, D. (2014). Non-rigid image registration: theory and practice.

Curtin, F. and Schulz, P. (1998). Multiple correlations and bonferronis correction. *Biological psychiatry*, 44(8):775–777.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.

Daugman, J. (2004). How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30.

Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161.

David, H. A. and Gunnink, J. L. (1997). The paired t test under artificial pairing. *The American Statistician*, 51(1):9–12.

de Oliveira, F. S. S., de Carvalho Filho, A. O., Silva, A. C., de Paiva, A. C., and Gattass, M. (2015). Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM. *Computers in biology and medicine*, 57:42–53.

de Oliveira Martins, L., Silva, A. C., De Paiva, A. C., and Gattass, M. (2009). Detection of breast masses in mammogram images using growing neural gas algorithm and ripleys k function. *Journal of Signal Processing Systems*, 55(1-3):77–90.

Dheeba, J., Singh, N. A., and Selvi, S. T. (2014). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of Biomedical Informatics*, 49:45–52.

Diez, Y., Oliver, A., Lladó, X., Freixenet, J., Marti, J., Vilanova, J. C., and Marti, R. (2011). Revisiting intensity-based image registration applied to mammography. *IEEE Transactions on Information Technology in Biomedicine*, 15(5):716–725.

Ding, J., Cheng, H., Huang, J., Liu, J., and Zhang, Y. (2012). Breast ultrasound image classification based on multiple-instance learning. *Journal of digital imaging*, 25(5):620–627.

Elad, M. and Hel-Or, Y. (2001). A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. *Image Processing, IEEE Transactions on*, 10(8):1187–1193.

Elmore, J. G., Armstrong, K., Lehman, C. D., and Fletcher, S. W. (2005). Screening for breast cancer. *Journal of the American Medical Association*, 293(10):1245–1256.

Farsiu, S., Robinson, M. D., Elad, M., and Milanfar, P. (2004). Fast and robust multiframe super resolution. *IEEE Transactions on Image processing*, 13(10):1327–1344.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874.

Fischer, B. and Modersitzki, J. (2004). A unified approach to fast image registration and a new curvature based registration technique. *Linear Algebra and its Applications*, 380:107–124.

Flores, W. G., de Albuquerque Pereira, W. C., and Infantosi, A. F. C. (2015). Improving classification performance of breast lesions on ultrasonography. *Pattern Recognition*, 48(4):1125–1136.

Gamdonkar, Z., Tay, K., Ryder, W., Brennan, P. C., and Mello-Thoms, C. (2015). idensity: an automatic Gabor filter-based algorithm for breast density assessment. In *SPIE Medical Imaging*, pages 941607–941607. International Society for Optics and Photonics.

García-Manso, A., García-Orellana, C., González-Velasco, H., Gallardo-Caballero, R., and Macías-Macías, M. (2013). Study of the effect of breast tissue density on detection of masses in mammograms. *Computational and mathematical methods in medicine*, 2013.

Gilbert, F. J., Astley, S. M., Gillan, M. G., Agbaje, O. F., Wallis, M. G., James, J., Boggis, C. R., and Duffy, S. W. (2008). Single reading with computer-aided detection for screening mammography. *New England Journal of Medicine*, 359(16):1675–1684.

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

Gómez, W., Pereira, W., and Infantosi, A. F. C. (2012). Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. *IEEE Transactions on Medical Imaging*, 31(10):1889–1899.

Gonzalez, R. C. and Woods, R. E. (2002). Digital image processing.

Guo, Y., Sivaramakrishna, R., Lu, C.-C., Suri, J. S., and Laxminarayan, S. (2006). Breast image registration techniques: a survey. *Medical and Biological Engineering and Computing*, 44(1-2):15–26.

Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, (6):610–621.

He, X.-C. and Yung, N. H. (2004). Curvature scale space corner detector with adaptive threshold and dynamic region of support. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 2, pages 791–794. IEEE.

Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr, P., Moore, R., Chang, K., and Munishkumaran, S. (1998). Current status of the digital database for screening mammography. In *Digital mammography*, pages 457–460. Springer.

Heikkilä, M., Pietikäinen, M., and Schmid, C. (2009). Description of interest regions with local binary patterns. *Pattern Recognition*, 42(3):425–436.

Hong, B.-W. and Brady, M. (2003). A topographic representation for mammogram segmentation. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*, pages 730–737. Springer.

Hong, B.-W. and Brady, M. (2005). Structural comparison of mammograms. In *British Machine Vision Conference*.

Hong, B.-W. and Sohn, B.-S. (2010). Segmentation of regions of interest in mammograms in a topographic approach. *Information Technology in Biomedicine, IEEE Transactions on*, 14(1):129–139.

Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–203.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.

Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. (2003). A practical guide to support vector classification.

Hsu, C.-W. and Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2):415–425.

Iakovidis, D. K., Keramidas, E. G., and Maroulis, D. (2008). Fuzzy local binary patterns for ultrasound texture characterization. In *Image Analysis and Recognition*, pages 750–759. Springer.

Jabid, T., Kabir, M. H., and Chae, O. (2010). Facial expression recognition using local directional pattern (ldp). In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1605–1608. IEEE.

Jalalian, A., Mashohor, S. B., Mahmud, H. R., Saripan, M. I. B., Ramli, A. R. B., and Karasfi, B. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clinical imaging*, 37(3):420–426.

Jas, M., Mukhopadhyay, S., Chakraborty, J., Sadhu, A., and Khandelwal, N. (2013). A heuristic approach to automated nipple detection in digital mammograms. *Journal of Digital Imaging*, 26(5):932–940.

Jen, C.-C. and Yu, S.-S. (2015). Automatic detection of abnormal mammograms in mammographic images. *Expert Systems with Applications*, 42(6):3048–3055.

Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library.

Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1233–1258.

Juszczak, P., Tax, D., and Duin, R. (2002). Feature scaling in support vector data description. In *Proc. ASCI*, pages 95–102.

Karssemeijer, N. (1998). Automated classification of parenchymal patterns in mammograms. *Physics in Medicine and Biology*, 43(2):365.

Kawahara, J., Peyrat, J.-M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Abugharbieh, R., and Hamarneh, G. (2014). Automatic labelling of tumourous frames in free-hand laparoscopic ultrasound video. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 676–683. Springer.

Keramidas, E., Iakovidis, D., and Maroulis, D. (2011). Fuzzy binary patterns for uncertainty-aware texture representation. *Electronic Letters on Computer Vision and Image Analysis*, 10(1):63–78.

Kirsch, R. A. (1971). Computer determination of the constituent structure of biological images. *Computers and Biomedical Research*, 4(3):315–328.

Koay, J., Herry, C., and Frize, M. (2004). Analysis of breast thermography with an artificial neural network. In *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 1159–1162. IEEE.

Krawczyk, B., Schaefer, G., and Woźniak, M. (2015). A hybrid cost-sensitive ensemble for imbalanced breast thermogram classification. *Artificial Intelligence in Medicine*, 65(3):219–227.

Krueger, M., Pesavento, A., Ermert, H., Hiltawsky, K., Heuser, L., Rosenthal, H., and Jensen, A. (1998). Ultrasonic strain imaging of the female breast using phase root seeking and three-dimensional optical flow. In *Proceedings of IEEE Ultrasonics Symposium.*, volume 2, pages 1757–1760. IEEE.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Lee, K. A. (1997). A mammographic registration method based on optical flow and multiresolution computing. Technical report, DTIC Document.

Liu, X. and Zeng, Z. (2015). A new automatic mass detection method for breast cancer with false positive reduction. *Neurocomputing*, 152:388–402.

Liu, Y. and Zheng, Y. F. (2005). One-against-all multi-class svm classification using reliability measures. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 2, pages 849–854. IEEE.

Lo, C., Chang, R., Huang, C., and Moon, W. (2015a). Computer-aided diagnosis of breast tumors using textures from intensity transformed sonographic images. In *1st Global Conference on Biomedical Engineering & 9th Asian-Pacific Conference on Medical and Biological Engineering*, pages 124–127. Springer.

Lo, C.-M., Moon, W. K., Huang, C.-S., Chen, J.-H., Yang, M.-C., and Chang, R.-F. (2015b). Intensity-invariant texture analysis for classification of BI-RADS category 3 breast masses. *Ultrasound in medicine & biology*, 41(7):2039–2048.

Lokate, M., Kallenberg, M. G., Karssemeijer, N., Van den Bosch, M. A., Peeters, P. H., and Van Gils, C. H. (2010). Volumetric breast density from full-field digital mammograms and its association with breast cancer risk factors: a comparison with a threshold method. *Cancer Epidemiology Biomarkers & Prevention*, 19(12):3096–3105.

Lombaert, H., Grady, L., Pennec, X., Ayache, N., and Cheriet, F. (2014). Spectral log-demons: Diffeomorphic image registration with very large deformations. *International Journal of Computer Vision*, 107(3):254–271.

Lowry, R. (2014). Concepts and applications of inferential statistics.

Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *In Proc. Seventh International Joint Conference on Artificial Intelligence*, volume 81, pages 674–679.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198.

Maitra, I. K., Nag, S., and Bandyopadhyay, S. K. (2011). Accurate breast contour detection algorithms in digital mammogram. *International Journal of Computer Applications*, 25(5):1–13.

Malvezzi, M., Bertuccio, P., Rosso, T., Rota, M., Levi, F., La Vecchia, C., and Negri, E. (2015). European cancer mortality predictions for the year 2015: does lung cancer have the highest death rate in eu women. *Ann Oncol*, 26(4):779–786.

Mandelbrot, B. B. (1983). *The fractal geometry of nature*, volume 173. Macmillan.

Marias, K., Behrenbruch, C., Brady, M., Parbhoo, S., and Seifalian, A. (2000). Multi-scale landmark selection for improved registration of temporal mammograms. In *Proc. International Workshop on Digital Mammography*, pages 580–586.

Marias, K., Behrenbruch, C., Parbhoo, S., Seifalian, A., and Brady, M. (2005). A registration framework for the comparison of mammogram sequences. *IEEE Transactions on Medical Imaging*, 24(6):782–790.

Marias, K., Brady, J., Highnam, R., Parbhoo, S., Seifalian, A., and Wirth, M. (1999). Registration and matching of temporal mammograms for detecting abnormalities. *Medical Imaging Understanding and Analysis*.

Melendez, J., Sánchez, C. I., van Ginneken, B., and Karssemeijer, N. (2014). Improving mass candidate detection in mammograms via feature maxima propagation and local feature selection. *Medical Physics*, 41(8):081904.

Mendelson, E., Böhm-Vélez, M., Berg, W., et al. (2013). Acr bi-rads® ultrasound. *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology*, pages 35–131.

Moghaddam, M. J., Tan, T., Karssemeijer, N., and Platel, B. (2014). Automatic nipple detection on 3d images of an automated breast ultrasound system (abus). In *SPIE Medical Imaging*, pages 903405–903405. International Society for Optics and Photonics.

Mohamed, M., Rashwan, H., Mertsching, B., Garcia, M., and Puig, D. (2014). Illumination-robust optical flow using local directional pattern.

Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., and Cardoso, J. S. (2012). Inbreast: toward a full-field digital mammographic database. *Academic radiology*, 19(2):236–248.

Muhimmah, I. and Zwiggelaar, R. (2006). Mammographic density classification using multiresolution histogram information. In *Proceedings of the International Special Topic Conference on Information Technology in Biomedicine, ITAB*.

Müller, T., Rabe, C., Rannacher, J., Franke, U., and Mester, R. (2011). Illumination-robust dense optical flow using census signatures. In *Pattern Recognition*, pages 236–245. Springer.

Muštra, M., Grgić, M., and Delač, K. (2012). Breast density classification using multiple feature selection. *AUTOMATIKA: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 53(4):362–372.

Ng, E. Y. and Sudharsan, N. (2004). Computer simulation in conjunction with medical thermography as an adjunct tool for early detection of breast cancer. *BMC cancer*, 4(1):17.

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987.

Oliveira, F. P. and Tavares, J. M. R. (2014). Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93.

Oliver, A., Freixenet, J., Marti, J., Perez, E., Pont, J., Denton, E. R., and Zwiggelaar, R. (2010). A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis*, 14(2):87–110.

Oliver, A., Freixenet, J., Marti, R., Pont, J., Pérez, E., Denton, E. R., and Zwiggelaar, R. (2008). A novel breast tissue density classification methodology. *IEEE Transactions on Information Technology in Biomedicine*, 12(1):55–65.

Oliver, A., Freixenet, J., and Zwiggelaar, R. (2005). Automatic classification of breast density. In *IEEE International Conference on Image Processing, 2005. ICIP 2005.*, volume 2, pages II–1258. IEEE.

Oliver, A., Lladó, X., Freixenet, J., and Martí, J. (2007). False positive reduction in mammographic mass detection using local binary patterns. In *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2007*, pages 286–293. Springer.

Oliver, A., Tortajada, M., Lladó, X., Freixenet, J., Ganau, S., Tortajada, L., Vilagran, M., Sentís, M., and Martí, R. (2015). Breast density analysis using an automatic density segmentation algorithm. *Journal of digital imaging*, pages 1–9.

Orel, S. G., Kay, N., Reynolds, C., and Sullivan, D. C. (1999). Bi-rads categorization as a predictor of malignancy 1. *Radiology*, 211(3):845–850.

Ou, Y., Sotiras, A., Paragios, N., and Davatzikos, C. (2011). Dramms: Deformable registration via attribute matching and mutual-saliency weighting. *Medical Image Analysis*, 15(4):622–639.

Ou, Y., Weinstein, S. P., Conant, E. F., Englander, S., Da, X., Gaonkar, B., Hsieh, M.-K., Rosen, M., DeMichele, A., Davatzikos, C., et al. (2015). Deformable

registration for quantifying longitudinal tumor changes during neoadjuvant chemotherapy. *Magnetic Resonance in Medicine*, 73(6):2343–2356.

Pereira, S. M. P., Hipwell, J. H., McCormack, V. A., Tanner, C., Moss, S. M., Wilkinson, L. S., Khoo, L. A., Pagliari, C., Skippage, P. L., Kliger, C. J., et al. (2010). Automated registration of diagnostic to prediagnostic x-ray mammograms: evaluation and comparison to radiologists? accuracy. *Medical Physics*, 37(9):4530–4539.

Pertuz, S., Julia, C., and Puig, D. (2014). A novel mammography image representation framework with application to image registration. In *22nd International Conference on Pattern Recognition (ICPR)*, pages 3292–3297. IEEE.

Petroudi, S., Constantinou, I., Pattichis, M., Tziakouri, C., Marias, K., and Pattichis, C. (2015). Evaluation of spatial dependence matrices on multiscale instantaneous amplitude for mammogram classification. In *6th European Conference of the International Federation for Medical and Biological Engineering*, pages 156–159. Springer.

Pisano, E. D., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J. K., Acharyya, S., Conant, E. F., Fajardo, L. L., Bassett, L., D'Orsi, C., et al. (2005). Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, 353(17):1773–1783.

Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., Braeuning, M. P., and Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital Imaging*, 11(4):193–200.

Pomponiu, V., Hariharan, H., Zheng, B., and Gur, D. (2014). Improving breast mass detection using histogram of oriented gradients. In *SPIE Medical Imaging*, pages 90351R–90351R. International Society for Optics and Photonics.

Puff, D. T., Pisano, E. D., Muller, K. E., Johnston, R. E., Hemminger, B. M., Burbeck, C. A., McLelland, R., and Pizer, S. M. (1994). A method for

determination of optimal image enhancement for the detection of mammographic abnormalities. *Journal of Digital Imaging*, 7(4):161–171.

Qian, W., Li, L., Clarke, L., Clark, R. A., and Thomas, J. (1999). Digital mammography: comparison of adaptive and nonadaptive CAD methods for mass detection. *Academic radiology*, 6(8):471–480.

Quan, Y., Xu, Y., Sun, Y., and Luo, Y. (2014). Lacunarity analysis on image patterns for texture classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 160–167. IEEE.

Rajagopal, V., Lee, A., Chung, J.-H., Warren, R., Highnam, R. P., Nash, M. P., and Nielsen, P. M. (2008). Creating individual-specific biomechanical models of the breast for medical image analysis. *Academic Radiology*, 15(11):1425–1436.

Ramirez Rivera, A., Castillo, R., and Chae, O. (2013). Local directional number pattern for face analysis: Face and expression recognition. *IEEE Transactions on Image Processing*, 22(5):1740–1752.

Rangayyan, R. M., Nguyen, T. M., Ayres, F. J., and Nandi, A. K. (2010). Effect of pixel resolution on texture features of breast masses in mammograms. *Journal of Digital Imaging*, 23(5):547–553.

Rashwan, H. A., Mohamed, M. A., García, M. A., Mertsching, B., and Puig, D. (2013). Illumination robust optical flow model based on histogram of oriented gradients. In *Pattern recognition*, pages 354–363. Springer.

Richard, F. J. and Cohen, L. D. (2003). A new image registration technique with free boundary constraints: application to mammography. *Computer Vision and Image Understanding*, 89(2):166–196.

Rohlfing, T., Maurer Jr, C. R., Bluemke, D., Jacobs, M., et al. (2003). Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging*, 22(6):730–741.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721.

Sallam, M. Y. and Bowyer, K. W. (1999). Registration and difference analysis of corresponding mammogram images. *Medical Image Analysis*, 3(2):103–118.

Sampat, M. P., Bovik, A. C., Whitman, G. J., and Markey, M. K. (2008). A model-based framework for the detection of spiculated masses on mammography. *Medical Physics*, 35(5):2110–2123.

Saniei, E., Setayeshi, S., Akbari, M. E., and Navid, M. (2015). A vascular network matching in dynamic thermography for breast cancer detection. *Quantitative InfraRed Thermography Journal*, (ahead-of-print):1–13.

Schaefer, G., Závišek, M., and Nakashima, T. (2009). Thermography based breast cancer analysis using statistical features and fuzzy classification. *Pattern Recognition*, 42(6):1133–1137.

Scheel, J. R., Lee, J. M., Sprague, B. L., Lee, C. I., and Lehman, C. D. (2015). Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. *American journal of obstetrics and gynecology*, 212(1):9–17.

Scholkopft, B. and Mullert, K.-R. (1999). Fisher discriminant analysis with kernels. *IX Neural networks for signal processing.*

Sehmbi, H. and Perlas, A. (2015). Basics of ultrasound imaging. In *Regional Nerve Blocks in Anesthesia and Pain Therapy*, pages 27–56. Springer.

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

Sharma, V. and Singh, S. (2014). Cfs–smo based classification of breast density using multiple texture models. *Medical & biological engineering & computing*, 52(6):521–529.

Sharma, V. and Singh, S. (2015). Automated classification of fatty and dense mammograms. *Journal of Medical Imaging and Health Informatics*, 5(3):520–526.

Shi, X., Cheng, H., Hu, L., Ju, W., and Tian, J. (2010). Detection and classification of masses in breast ultrasound images. *Digital signal processing*, 20(3):824–836.

Siegel, R. L., Miller, K. D., and Jemal, A. (2015). Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians*, 65(1):5–29.

Silva, L., Saade, D., Sequeiros, G., Silva, A., Paiva, A., Bravo, R., and Conci, A. (2014). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1):92–100.

Singh, S. and Al-Mansoori, R. (2000). Identification of regions of interest in digital mammograms. *Journal of Intelligent Systems*, 10(2):183–217.

Soh, L.-K. and Tsatsoulis, C. (1999). Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):780–795.

Soltanian-Zadeh, H., Rafiee-Rad, F., and Pourabdollah-Nejad D, S. (2004). Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms. *Pattern Recognition*, 37(10):1973–1986.

Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190.

Štruc, V. and Pavešić, N. (2009). Phase congruency features for palm-print verification. *IET Signal Processing*, 3(4):258–268.

Subashini, T., Ramalingam, V., and Palanivel, S. (2010). Automated assessment of breast tissue density in digital mammograms. *Computer Vision and Image Understanding*, 114(1):33–43.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S.-L., et al. (1994). The mammographic image analysis society digital mammogram database. In *Proc. the 2nd International Workshop on Digital Mammography, Amsterdam, Netherlands*, pages 375–378.

Sun, D., Roth, S., and Black, M. J. (2010). Secrets of optical flow estimation and their principles. In *23rd IEEE Conference on Computer Vision and Pattern Recognition, CVPR, San Francisco, CA, USA, 13-18 June*, pages 2432–2439.

Sun, T. and Neuvo, Y. (1994). Detail-preserving median based filters in image processing. *Pattern Recognition Letters*, 15(4):341–347.

Teo, P. and Pistorius, S. (2014). Tissue motion tracking at the edges of a radiation treatment field using local optical flow analysis. In *Journal of Physics Conference Series*, volume 489, page 2040.

Tortajada, M., Oliver, A., Martí, R., Vilagran, M., Ganau, S., Tortajada, L., Sentís, M., and Freixenet, J. (2012). Adapting breast density classification from digitized to full-field digital mammograms. In *Breast Imaging*, pages 561–568. Springer.

Uniyal, N., Eskandari, H., Abolmaesumi, P., Sojoudi, S., Gordon, P., Warren, L., Rohling, R. N., Salcudean, S. E., and Moradi, M. (2015). Ultrasound rf time series for classification of breast lesions. *IEEE Transactions on Medical Imaging*, 34(2):652–661.

van Engeland, S., Snoeren, P., Hendriks, J., and Karssemeijer, N. (2003). A comparison of methods for mammogram registration. *IEEE Transactions on Medical Imaging*, 22(11):1436–1444.

Veeramani, S. K. and Muthusamy, E. (2015). Detection of abnormalities in

ultrasound lung image using multi-level rvm classification. *The Journal of Maternal-Fetal & Neonatal Medicine*, pages 1–9.

Wai, L. C. and Brady, M. (2005). Curvilinear structure based mammographic registration. In *Computer Vision for Biomedical Image Applications*, pages 261–270. Springer.

Wang, L., Böhler, T., Zöhrer, F., Georgii, J., Rauh, C., Fasching, P. A., Brehm, B., Schulz-Wendtland, R., Beckmann, M. W., Uder, M., et al. (2014a). Fully automated nipple detection in 3d breast ultrasound images. In *Breast Imaging*, pages 64–71. Springer.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Wang, Z., Yu, G., Kang, Y., Zhao, Y., and Qu, Q. (2014b). Breast tumor detection in digital mammography based on extreme learning machine. *Neurocomputing*, 128:175–184.

Weldon, T. P., Higgins, W. E., and Dunn, D. F. (1996). Efficient Gabor filter design for texture segmentation. *Pattern Recognition*, 29(12):2005–2015.

Wolfe, J. N. (1976). Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, 37(5):2486–2492.

Wu, J., Ou, Y., Weinstein, S. P., Conant, E. F., Yu, N., Hoshmand, V., Keller, B., Ashraf, A. B., Rosen, M., DeMichele, A., et al. (2015). Quantification of tumor changes during neoadjuvant chemotherapy with longitudinal breast dce-mri registration. In *SPIE Medical Imaging*, pages 94141Z–94141Z. International Society for Optics and Photonics.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190.

Yager, R. R. (1993). Families of owa operators. *Fuzzy sets and systems*, 59(2):125–148.

Yang, M.-C., Moon, W. K., Wang, Y.-C. F., Bae, M. S., Huang, C.-S., Chen, J.-H., and Chang, R.-F. (2013). Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis. *IEEE Transactions on Medical Imaging*, 32(12):2262–2273.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3):338–353.

Zheng, Y. (2010). Breast cancer detection with Gabor features from digital mammograms. *algorithms*, 3(1):44–62.

Zheng, Y., Keller, B. M., Ray, S., Wang, Y., Conant, E. F., Gee, J. C., and Kontos, D. (2015). Parenchymal texture analysis in digital mammography: A fully automated pipeline for breast cancer risk assessment. *Medical Physics*, 42(7):4149–4160.

Zhou, Z., Wu, S., Chang, K.-J., Chen, W.-R., Chen, Y.-S., Kuo, W.-H., Lin, C.-C., and Tsui, P.-H. (2015). Classification of benign and malignant breast tumors in ultrasound images with posterior acoustic shadowing using half-contour features. *Journal of medical and biological engineering*, 35(2):178–187.

Zwiggelaar, R. (2010). Local greylevel appearance histogram based texture segmentation. In *Digital Mammography*, pages 175–182. Springer.

Acronyms

The acronyms are arranged in the order of the first appearance in the work.

CAD computer-aided diagnosis, page 3

ROIs regions of interest, page 4

ULDP uniform local directional pattern, page 6

FLDP fuzzy local directional pattern, page 7

ROI region of interest, page 10

CC cranio-caudal, page 14

MLO medioLateral oblique, page 14

BUS breast ultrasonography, page 14

MRI magnetic resonance imaging, page 16

GT ground truth, page 16

LBP local binary pattern, page 17

HOG histogram of oriented gradients, page 17

GLCM grey level co-occurrence matrix, page 17

k-NN k-nearest neighbor, page 21

LDA linear discriminant analysis, page 21

SVM support vector machine, page 22

LSVM linear support vector machine, page 22

NLSVM nonlinear support vector machine, page 22

RF random forests, page 23

MLP multi-layer perceptron, page 23

TP true positive, page 24

TN true negative, page 24

FP false positive, page 24

FN false negative, page 24

AUC area under the curve, page 24

ROC receiver operating characteristic, page 24

LOOCV leave-one-out-cross validation, page 25

MI mutual information, page 28

CTF coarse-to-fine technique, page 29

LDN local directional number patterns, page 36

RLBP robust local binary pattern, page 40

FLBP fuzzy local binary pattern, page 40

CSLBP center symmetric local binary pattern, page 40

LGA local grey level appearance, page 43

PCA principal component analysis, page 48

MLDP modified local directional pattern, page 59

CLAHE contrast-limited adaptive histogram equalization, page 81

SH sharpening, page 81

MF median filter, page 81

HAR Haralick's features, page 81

GF Gabor filters, page 81

SFS sequential forward selection, page 82

ExS exhaustive search, page 82

zs zero mean unit variance data normalization, page 85

mn maximum-minimum data normalization, page 85

ANOVA analysis of variance, page 86

Res pixel resolution, page 88

IS integration scale, page 88

NP no preprocessing, page 89

PCLBP phase congruency with local binary pattern, page 114

SR super-resolution, page 115

HR high-resolution, page 115

LR low-resolution, page 115

OIS one image system, page 124

HE histogram equalization, page 124

HBB human body boundary, page 138

FFDM full field digital mammograms, page 156

DICOM digital imaging and communication in medicine, page 156

SSIM structural similarity, page 167

OSF oversampling factor, page 168

LK Lucas-Kanade optical flow method, page 178

OWA ordered weighted averaging aggregation, page 178

HS Horn−Schunck optical flow method, page 181

LDOF large displacement optical flow, page 181

GNC graduated non-convexity, page 189