University of Plymouth

PEARL **https://pearl.plymouth.ac.uk**

04 University of Plymouth Research Theses 01 Research Theses Main Collection

2004

Analysis of Dynamic Magnetic Resonance Breast Images

de Pasquale, Francesco

http://hdl.handle.net/10026.1/2351

http://dx.doi.org/10.24382/4688
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

UNIVERSITY OF PLYMOUTH

Analysis of Dynamic Magnetic Resonance Breast Images

by

Francesco de Pasquale

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

Doctor of Philosophy

School of Mathematics and Statistics, Faculty of Technology,
21st February 2004

UNIVERSITY OF PLYMOUTH
Item No. 9005878460
Date 05 MAY 2004
Class No. THESIS 614.8465 DE
Cont. No.
PLYMOUTH LIBRARY

A BRIEF WORD OF EXPLANATION
OUR ASSIGNMENT WAS A TWO-THOUSAND WORD REPORT
I HAVE HEARD IT SAID THAT ONE PICTURE IS WORTH A THOUSAND WORDS...
WHAT WE HAVE HERE IS A COUPLE OF PICTURES...

Analysis of Dynamic Magnetic Resonance Breast Images

Abstract

Francesco de Pasquale

Dynamic Magnetic Resonance Imaging is a non-invasive technique that provides an image sequence based on dynamic information for locating lesions and investigating their structures.

In this thesis we develop new methodology for analysing dynamic Magnetic Resonance image sequences of the breast. This methodology comprises an image restoration step that reduces random distortions affecting the data and an image classification step that identifies normal, benign or malignant tumoral tissues.

In the first part of this thesis we present a non-parametric and a parametric approach for image restoration and classification. Both methods are developed within the Bayesian framework. A prior distribution modelling both spatial homogeneity and temporal continuity between neighbouring image pixels is employed. Statistical inference is performed by means of a Metropolis-Hastings algorithm with a specially chosen proposal distribution that out-performs other algorithms of the same family. We also provide novel procedures for estimating the hyper-parameters of the prior models and the normalizing constant so making the Bayesian methodology automatic.

In the second part of this thesis we present new methodology for image classification based on deformable templates of a prototype shape. Our approach uses higher level knowledge about the tumour structure than the spatio-temporal prior distribution of our Bayesian methodology. The prototype shape is deformed to identify the structure of the malignant tumoral tissue by minimizing a novel objective function over the parameters of a set of non-affine transformations. Since these transformations can destroy the connectivity of the shape, we develop a new filter that restores connectivity without smoothing the shape.

The restoration and classification results obtained from a small sample of image sequences are very encouraging. In order to validate these results on a larger sample, in the last part of the thesis we present a user friendly software package that implements our methodology.

Contents

Abstract i

List of Figures vi

List of Tables x

Author's declaration xi

1 Introduction **1**

1.1 Aims and outline of the thesis 1

1.2 Dynamic Magnetic Resonance Imaging of the breast 4

1.3 Bayesian image analysis 9

1.3.1 Image data model 10

1.3.2 Prior model 11

1.3.3 Markov Random Fields 12

1.3.4 Bayesian inference 13

1.3.5 Markov chain Monte Carlo simulations 14

1.4 Image Classification 18

1.4.1 K-means clustering algorithm 19

1.4.2 The Iterative Conditional Mode algorithm 20

1.5 Deformable template models 21

1.5.1 Free form models 23

1.5.2 Parametric deformable models 24

2 The Data **27**

2.1 The physics of Magnetic Resonance Images 27

2.1.1 Nuclear Magnetic Resonance principles 28

2.1.2 Magnetic Resonance Imaging 29

2.2 The analysed dataset 30

3 Bayesian Image Restoration **35**

3.1 Region Of Interest selection and tumoral region location 36

3.2 A non-parametric approach for image restoration 39

3.2.1 Image data model 39

3.2.2 Prior model 40

3.2.3 Estimation 41

3.2.4 Hyper-parameter estimation 45

3.3 A parametric approach for image restoration 47

3.3.1 Image data model 47

3.3.2 Prior model 48

3.3.3 Estimation 49

3.3.4 Hyper-parameter estimation 50

3.4 Results 51

3.4.1 Non-parametric approach 51

3.4.2 Parametric approach 56

3.5 Conclusions 58

4 Bayesian Image Classification **62**

4.1 Attribute model 64
4.2 Prior model 64
4.3 Estimation 64
4.4 Results 66
4.5 Conclusions 69

5 Deformable Template Models **71**
5.1 The prototype template 73
5.2 The deformation model 74
5.2.1 Filters to recover the connectivity 83
5.3 The richness of the deformation model 87
5.4 The objective function 90
5.5 The matching algorithm 96
5.6 Results 98
5.6.1 Synthetic data 99
5.6.2 Real data 109
5.7 Conclusions 111

6 BAnDITS: a Software Package for the Analysis of dMRI Sequences **114**

7 Discussion and Future Work **120**

A Estimating the Best Fitting Ellipse of a Shape from its Moments **124**

B Morphological Operations **127**
B.1 Skeletonisation 127
B.2 Bridge transformation 131

List of Figures

1.1 Typical dMRI uptake patterns for tumoural breast tissue pixels. 7

1.2 Common morphological signs for breast lesions. 8

2.1 MRI spatial encoding. 31

2.2 The acquired dMRI sequence. 33

2.3 Two images from the acquired dMRI sequence with the region Of interest. . 34

2.4 Random distortions in the Region Of Interest. 34

3.1 The 'mean difference image' $\mathbf{m}$, the selected ROI and the reference region $\mathcal{R}$. 37

3.2 The behaviour of the prior potential $V_s(\Delta y_s)$ corresponding to different values of δ_s. 42

3.3 MSE comparisons between Metropolis algorithms with different proposals . 44

3.4 The parameters $\boldsymbol{\theta} = (I, M, F, p, \tau)$ of the parametric temporal model for the contrast agent uptake. 49

3.5 Comparison between the empirical and estimated distributions of pixel values within $\mathcal{R}$. 51

3.6 Results of the hypothesis test for different significance levels. 52

3.7 (a) Comparison between $p_l(\Delta y_l)$ for the optimal δ_l and the empirical distribution $\hat{p}_l(\Delta y_l)$ for both space and time. (b) Comparison between the theoretical χ^2_T density and the empirical distribution of $\Sigma_{\mathcal{I}}$ for the optimal β_s and β_t. 53

3.8 Non-parametric restorations. 55

3.9 The restoration obtained using a 3D Wavelet filter. 56

3.10 The temporal pattern before and after the non-parametric restoration procedure for two pixels within the tumoral region. 57

3.11 Images of the estimated attributes τ and $M - F$ from the parametric approach in the tumoral region $\mathcal{I}$. 58

3.12 Parametric restoration. 58

3.13 Comparisons between the acquired image intensity time profiles and the restored profile from the non-parametric and the parametric approach. . . . 59

3.14 Point estimates and 95% credibility intervals for a central row of the parameter τ shown in Fig. 3.11. 60

4.1 Classification attributes for a given pixel. 63

4.2 Non-parametric classification attributes A and B. 66

4.3 (a) The lesion location based on the mean difference image $\mathbf{m}$. (b) Hierarchical clustering classification. (c) Non-parametric classification. (d) Parametric classification. 67

4.4 The mean temporal patterns over the two classes. 69

5.1 Shifting and scaling the contour supplied by the radiologist. 75

5.2 The reference grid $\mathbf{R}$. 79

5.3 The effect on the reference grid $\mathbf{R}$ of various deformations. 80

5.4 An example of a deformation obtained as a composition of three transformations. 83
5.5 An unconnected shape together with the results of applying different filters. 85
5.6 Visiting scheme for neighbouring pixels. 86
5.7 Non-parametric Bayesian classification using only the attributes **A**. 88
5.8 The malignant tumoral region and the 'true' shape to be reconstructed in the simulation. 89
5.9 Comparison between the 'true' shape and its reconstructions. 90
5.10 The interior, the exterior, the internal edge and the external edge of the shape **T** 91
5.11 Two cases in which it is difficult to find the internal pixels of the shape. . . 93
5.12 A flowchart representing the matching algorithm. 97
5.13 True shape obtained by deforming the prototype template. 100
5.14 Data generated from the true shape, and the reconstructed and true lesion contour. 100
5.15 Data generated from the true shape, and the reconstructed and true lesion contour. 102
5.16 The second true shape. 103
5.17 Data generated from the true shape, and the reconstructed and true lesion contour. 104
5.18 The reconstructed and true lesion contour obtained by the modified algorithm. 105
5.19 The value of the objective function after each iteration of the matching algorithm. 106
5.20 The locked and unlocked pixels at each iteration of the modified algorithm. 107

5.21 Data generated from the true shape, and the reconstructed and true lesion contour. . . . 108
5.22 Data generated from the true shape, and the reconstructed and true lesion contour. . . . 109
5.23 The classification attributes **A** together with the reference region, and the final dataset. . . . 111
5.24 The final lesion reconstruction. . . . 112

6.1 BAnDITS opening screen. . . . 115
6.2 The ROI selection step. . . . 117
6.3 Classification results. . . . 118

B.1 The composite structuring elements. . . . 129
B.2 The visiting order for neighbouring pixels. . . . 130
B.3 The effect of Skeletonisation. . . . 131
B.4 The effect of the Bridge transformation. . . . 132

List of Tables

3.1 Hyper-parameter optimization ranges and the optimal values obtained. . . . 52

5.1 Estimates of the distribution parameters based on the true and reconstructed shape from the data shown in Fig. 5.14(a). 101

5.2 Estimates of the distribution parameters based on the true and reconstructed shape from the data shown in Fig. 5.15(a). 102

5.3 Estimates of the distribution parameters based on the true and reconstructed shape from the data shown in Fig. 5.17(a). 104

5.4 Estimates of the distribution parameters based on the true and reconstructed shape obtained by the modified algorithm for the data shown in Fig. 5.17(a). 105

5.5 Estimates of the distribution parameters based on the true and reconstructed lesion obtained by applying the modified algorithm to the data shown in Fig. 5.21(a). 108

5.6 Estimates of the distribution parameters based on the true and reconstructed lesion obtained by applying the modified algorithm to the data shown in Fig. 5.22(a). 110

Author's declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.
During my PhD studies I spent twelve months working at the Istituto per le Applicazioni del Calcolo, Consiglio Nazionale delle Ricerche, Rome, Italy.

Conferences attended and presentations:

In September 2001 I attended the conference *Mathematical Methods and Models for Studying Biological Phenomena* at the National Research Council in Rome. During this conference I gave a talk entitled "**Analysis of dynamic Magnetic Resonance images for breast tumour study**". I also attended three one-day workshops of the Royal Statitical Society Image Analysis and Processing Group. These were held at the Universities of Bath, Bristol and Nottingham. At the Nottingham workshop I made a presentation entitled "**Mammographic image analysis**".

Publications:

- de Pasquale, F., Barone, P., Sebastiani, G. and Stander J. (2003) Bayesian analysis of dynamic Magnetic Resonance breast images. Accepted for publication in *J. Roy. Statist. Soc.* Ser. C.
- de Pasquale, F., Soldaini, R., Testa, C., Casieri, C., Podo, F. and de Luca, F. (2003) Bayesian analysis of in-vivo ^{13}C-^{1}H images characterized by a very low signal to noise ratio. Submitted to *J. Magn. Reson.*
- de Pasquale F., Luciani A.M., Pacilio M., Guidoni L., Viti V., D'Errico F., Barone P. and Sebastiani G. (2002) Dose reconstruction in irradiated Fricke-agarose gels by means of MRI and optical techniques: 2D modelling of diffusion of ferric ions. *Rad. Prot. Dosimetry*, **99**, 363–364.
- de Pasquale, F., Sebastiani, G., Egger, E., Guidoni, L., Luciani, A.M., Marzola, P., Manfredi, R., Pacilio, M., Piermattei, A., Viti, V. and Barone, P. (2000) Bayesian estimation of relaxation times T1 in MR images of irradiated Fricke-agarose gels. *Magn. Reson. Im.*, **18**, 721–731.

Signed F. de Pasquale
Date 2004

Chapter 1

Introduction

1.1 Aims and outline of the thesis

The aim of the work presented in this thesis is to develop a new integrated methodology for analyzing dynamic Magnetic Resonance images (dMRI) of the breast. Our objective is to improve the diagnostic capability of this technique for locating breast tumours and investigating their structure. The dMRI analysis consists of two steps, an image restoration step followed by an image classification step. First, we try to minimize the distortions affecting the data. Then we classify every pixel of the breast as belonging to one of the following three classes: normal, tumoral benign or tumoral malignant tissue.

The study presented in this thesis comprises two main parts. In the first part a Bayesian methodology is presented. In this framework a non-parametric and a parametric approach are developed for image restoration. In the non-parametric approach this task is performed without adopting any model for the signal temporal evolution at each pixel. On the other hand, in the parametric approach a specific model is adopted for the signal temporal evolution. The parameters of this model are estimated and based on these estimates the restored sequence can be obtained. Image classification is then based on attributes acquired from the image restoration step. In the non-parametric

approach these attributes are calculated from the restored dMRI sequences, while in the parametric approach the attributes are some parameters of the adopted temporal model. In these methods statistical inference is performed by means of Markov chain Monte Carlo algorithms. In particular, we present a Metropolis-Hastings algorithm with a specially chosen proposal distribution that performs better than more commonly used proposals. Furthermore, we present a method to identify tumoral regions inside the breast and we also provide procedures for hyper-parameter estimation. These make the Bayesian methodology automatic. Most of the work described in the first part of this thesis is summarized in de Pasquale *et al.* (2003).

In the second part of this thesis we present a new algorithm based on deformable templates for image segmentation and classification. Whilst in the Bayesian methodology the prior distributions model spatial homogeneity and temporal continuity between neighbouring pixels, in this approach we assume 'a priori' knowledge about the tumour contour. This is represented by a prototype shape. Our algorithm deforms this shape by means of parametric non-affine transformations and the lesion contour is reconstructed by minimizing a newly developed objective function that depends on the transformation parameters. We show reconstructions obtained from synthetic and real data.

The thesis is structured as follows. In the rest of this chapter we present a general discussion about dynamic Magnetic Resonance Imaging of the breast. This is followed by an overview of Bayesian image analysis, image classification techniques and deformable template models. In Chapter 2 we briefly introduce the physical basis of Magnetic Resonance Imaging and describe the acquired data and experimental set-up.
In Chapter 3 we describe our Bayesian methodology for image restoration. We begin by presenting a method for identifying a tumoral region inside the breast. In this way, we can restrict our analysis to this region thus reducing considerably the computational burden

of our procedure. Next, we describe our non-parametric and parametric approaches for image restoration. We present the restoration that we obtain together with a comparison of our method with a three dimensional Wavelet based filter. Chapter 3 ends with a brief discussion of our restoration results.

In Chapter 4 the Bayesian image classification procedure is presented. We compare classification results obtained from non-parametric and parametric attributes with the results previously obtained by physicians. We then make comparisons with the classifications achieved from a standard hierarchical clustering method. The last section of this chapter is dedicated to a brief discussion of the results.

In Chapter 5 the deformable template based method is introduced. First, we describe how a prototype shape is chosen. We then introduce the parametric non-affine deformation model. Since deformations from this model can sometimes destroy the connectivity of the prototype shape, we developed a new filter that restores connectivity without smoothing the shape. Next, we investigate the richness of the space of shapes that the non-affine deformation model can generate and we discuss our novel objective function. For this, we modify our new filter to identify the interior of a shape. Subsequently, we describe a new matching algorithm and present the results obtained by applying this algorithm to real and synthetic data. In the last section of this chapter we summarize the results obtained with our deformable template based method.

Chapter 6 contains a brief description of the software package, called BAnDITS, that we have developed, together with some details of its implementation.

Finally, in Chapter 7 an overall discussion of the work of this thesis is presented, together with plans for the future work.

1.2 Dynamic Magnetic Resonance Imaging of the breast

Breast cancer is the second largest cause of cancer mortality in women today with 380000 deaths registered worldwide in 2000 (Ferlay *et al.*, 2001). More than a million patients are diagnosed with this disease every year and approximately 10% of women will develop it during the course of their lives (Hayton *et al.*, 1999). In order to reduce the rate of mortality due to breast cancer, it is very important to locate the lesions at a very early stage and to investigate their structure to plan specific and swift clinical interventions. Different techniques such as X-ray mammography, biopsy (transcutaneous, core and needle) and thermography are well known to be effective for diagnosing breast cancer. Despite their diagnostic capability there is room for considerable improvement: currently more than one operation in two is performed to remove a lesion that turns out to be benign; 8–25% of cancers are missed and 70–80% of biopsies turn out to be benign (Hayton *et al.*, 1996). Accordingly, the results of the current diagnostic techniques are generally not very specific. To overcome these shortcomings, in the last ten years new imaging modalities complementary to X-ray mammography have been introduced. These include Ultrasonography (2D, 3D and various types of Doppler imaging), Positron Emission Tomography (PET), Single Photon Computed Tomography (SPECT) and contrast enhanced or dynamic Magnetic Resonance Imaging (dMRI). Nowadays, these imaging techniques play a crucial role in screening breast cancer. In fact, radiologists and technicians base their diagnosis mainly on the analysis of the acquired images. Thus, image processing is a key step in the decision making procedure.

Among these techniques, X-ray mammography is most appropriate for general screening (Highnam and Brady, 1999). Its main advantage over other imaging modalities

(such as dMRI) is that it is cheaper and quicker. In fact, a single image is acquired (not a sequence as in dMRI). Despite this advantage, X-ray mammography has three fundamental drawbacks: it is projective, it has a limited applicability and it cannot image the dynamics of the tumour. In fact, to reduce the radiation dosage that the patient receives, the breasts are tightly compressed between parallel plates while the mammogram is taken. Therefore, the most that a single mammogram can show is the integral of non-adipose tissue in the direction of the X-ray beam. It follows that localization of a lesion or a microcalcification is inherently poorer than might be hoped for with truly three-dimensional imaging. In addition, mammographic screening is limited in application to women over 50 years of age since its benefit for younger women has not yet been demonstrated. The major reason is that before involution the breast contains a great deal of glandular tissue that is radio-opaque. At involution this tissue changes to fat which is transradiant. Whenever they are used, mammograms provide an image of breast anatomy. However, malignant cancers are characterized more by their dynamic behaviour. To nourish its growth, a tumour sprouts a network of new blood vessels and this causes an increase in the local blood supply (angiogenesis), which mammography cannot show. For these limitations it is very important to develop new imaging techniques complementary to X-ray mammography, particularly for younger women and especially for those known from epidemiological studies to be at high risk of developing breast cancer.

Amongst the techniques developed in the last ten years, dMRI seems one of the most promising. In contrast to X-ray based techniques, with dMRI the patient absorbs energy of ten orders of magnitude smaller ($10^{21}\nu$ for X-ray compared to $10^{9}\nu$ for MRI, where ν is the radiation frequency). In addition, the breast can be imaged in 3-D or 2-D and dynamic information about angiogenesis can be collected. The new information provided by dMRI appears to be very promising in the following situations (Heywang-Kobrunner

and Beck, 1995) :

- severe scarring after surgery with or without radiation therapy
- exclusion or early detection of malignancy around/behind silicon implants
- evaluation of implant integrity
- dense breast
- pre-operative examinations for breast conserving surgery
- patients with occult primary breast carcinoma
- monitoring of high risk patients (inherited mutation of the BRCA1 or BRCA2 genes).

dMRI consists of the acquisition of a sequence of MR images of a given section of the breast after the injection of a contrast agent into the blood stream. Usually a gadolinium salt (Gd-DTPA) is used as a contrast agent (Heywang *et al.*, 1986). The Gd-DTPA distributes itself in the intra-vascular or interstitial spaces of the breast. The consequent enhancement that occurs within malignancies is believed to be due to a variable combination of increased vascularity, increased permeability and/or increased interstitial space in malignant tumors compared with benign ones. The local concentration of the contrast agent modifies the MR image intensities and different breast tumoral tissues show a typical uptake pattern over the time (Hayton *et al.*, 1999). In Fig.1.1 the typical uptake patterns for the tumoral tissues are shown. Important features of these patterns are the speed of initial intensity variation and the presence of a final decrease. In fact, these features can be connected to contrast agent wash-in and wash-out properties. Because the vast majority (around 98%) of invasive malignancies enhance, dMRI is a highly specific technique unlike other form of mammography.

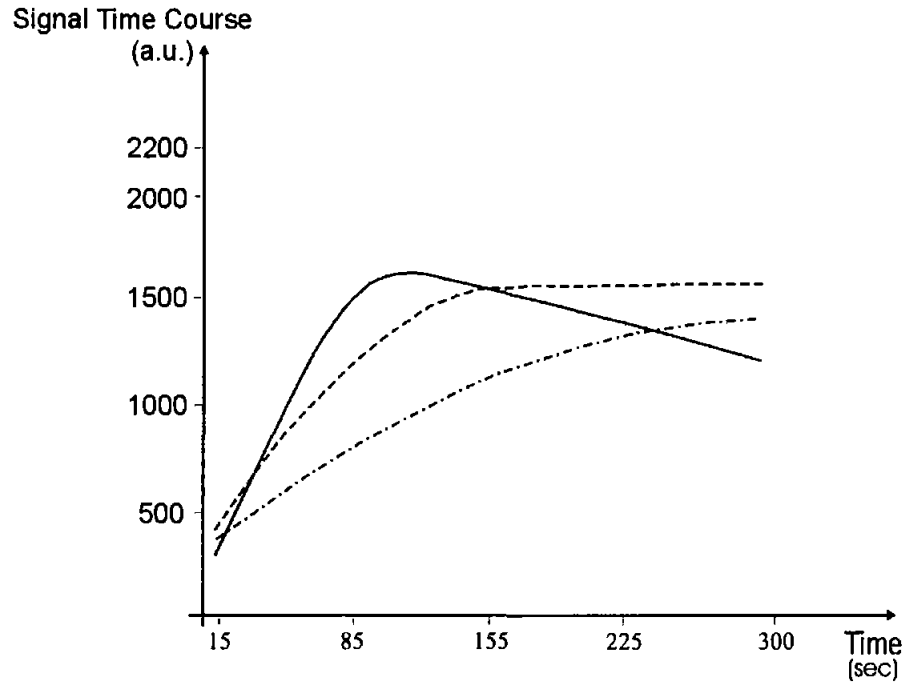


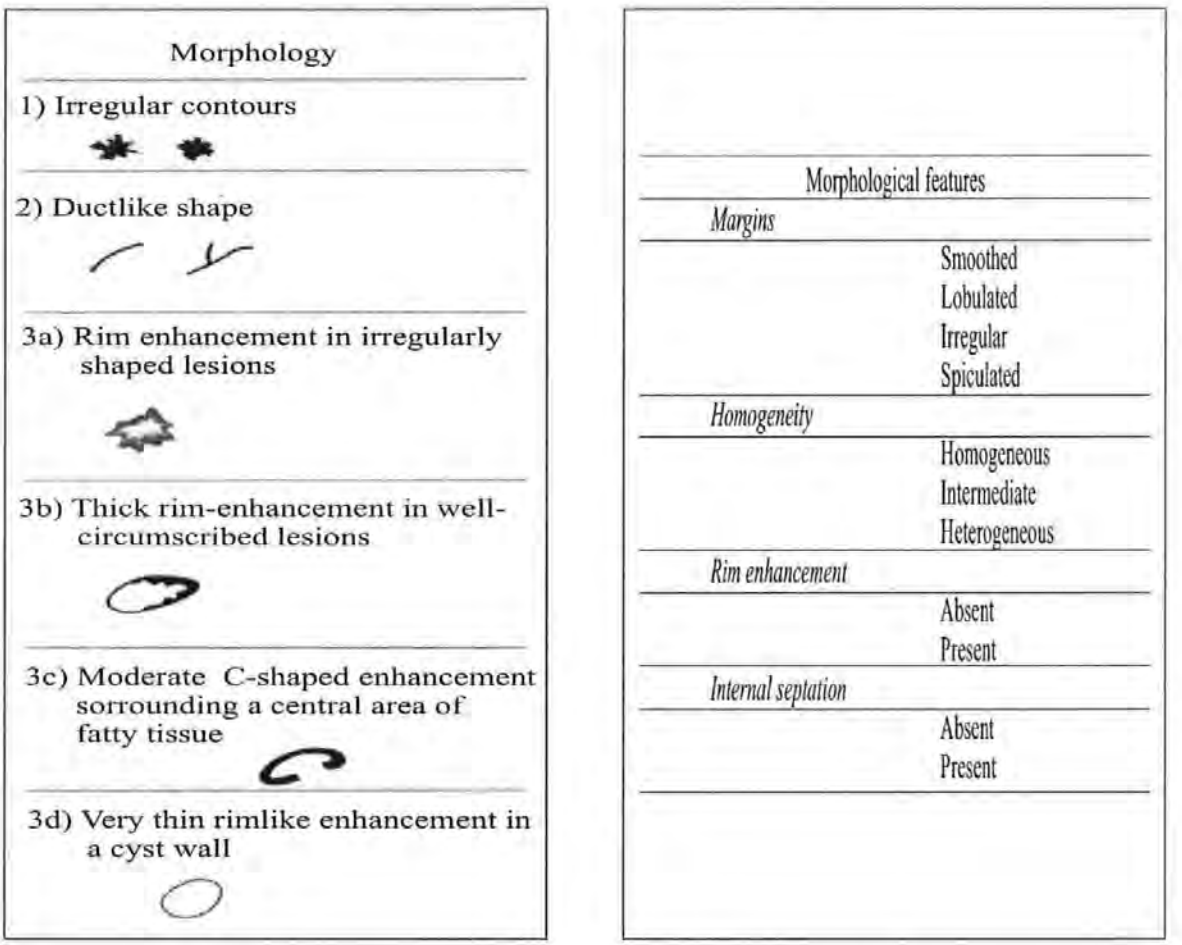
Figure 1.1: *Typical dMRI uptake patterns for tumoural breast tissue pixels. The curve showing a clear maximum is often associated with malignant lesions. The other two curves are mostly associated with benign lesions. a.u. stands for arbitrary units.*

The following features are usually taken into account for the interpretation of dMRI images:

- presence of enhancement
- amount of the enhancement
- speed of enhancement and washout
- morphology of an enhanced region.

In particular, the lesion morphology (irregularly shaped, well circumscribed, lobulated, oval or round lesions) represents very important information for classifying tumours, as shown in Fig. 1.2. In this figure we report some of the morphological features taken into

account and morphological diagnostic criteria by Szabó *et al.* (2003).



(a) (b)

Figure 1.2: *(a) Some of the most common morphological signs for breast lesions. The lesion morphology represents very important information for classifying tumours (Heywang-Kobrunner, 1995). (b) As an example we report some of the morphological features taken into account by Szabó et al. (2003) in classifying these kinds of lesions.*

Furthermore, in some lesions an internal structure may be visible and diagnostically useful. This structure can represent two different patho-physiological conditions of the tumoral tissue, that in the following we will indicate for simplicity as malignant and benign tumoral tissues. Hence, to localize the lesion, to highlight its morphology and to

investigate its structure are the main goals of breast image analysis. These tasks can be better achieved by using the specific information about contrast agent uptake properties provided by dMRI data.

There has been much research about the analysis of dMRI data. Important references are Kuhl *et al.* (1999), Mussurakis *et al.* (1997, 1998) and Gribbestad *et al.* (1992), although these authors make no attempt to remove either deterministic distortions due to patient movement during the image acquisition or random distortions affecting the data. Hayton *et al.* (1999) and Krishnan *et al.* (1999) present methodology to correct for breast motion, while Hayton *et al.* (1996) use a pharmacokinetic model for the acquired signals to localize tumours. The '3TP method' of Weinstein *et al.* (1999) uses three images of the sequence to identify certain pathophysiological features.

There is now a huge literature concerned with removing random distortions from images. Good reviews of the most popular techniques such as Wavelets, Splines and linear filters are given by Abramovich *et al.* (2000). A seminal reference for the Bayesian approach that we adopt is Geman and Geman (1984).

1.3 Bayesian image analysis

Bayesian image analysis makes use of explicit probability models to incorporate general and scene-specific prior knowledge into the processing of images. It provides a unified framework within which many different image analysis tasks can be considered (Winkler, 1995). In this approach two types of information can be taken into account: empirical data and prior knowledge. The former is the dataset that we acquired from our experiment, while the latter represents the information that is available about the data before the experiment is performed. Hence, there are two sources of uncertainty or randomness that are considered in this framework. In fact, the empirical data are a degraded version of

ideal data and the prior knowledge is usually incomplete.

One of the main features of the Bayesian approach is to exploit the available prior knowledge. This allows us to reduce efficiently the distortions that usually affect image data. Since both the restoration and classification methods presented in Chapter 3 and 4 are developed within the Bayesian framework, in the following we describe this approach in detail. Two key ingredients are the image data model and the prior model.

1.3.1 Image data model

In order to describe the Bayesian paradigm we need to introduce some notation. We denote by $\mathbf{y}$ and $\mathbf{x}$ the acquired and 'true' images respectively. In this thesis every $m \times m$ image will be indicated by an array such that $\mathbf{x} = (x_s)_{s \in S}$ and $\mathbf{y} = (y_s)_{s \in S}$ where S is a square lattice whose points represent the image pixels so that $|S| = m^2 = n$.

Formally, distortions affecting the acquired data $\mathbf{y}$ can be either deterministic or random, so that the most general relationship between $\mathbf{y}$ and $\mathbf{x}$ can be represented as:

$$\mathbf{y} = \mathcal{B}(\mathbf{x}) + \epsilon$$

where $\mathcal{B}$ is a 'Blur operator' and ϵ is a random variable. In general $\mathcal{B}$ is highly non-invertible and represents deterministic distortions, while ϵ represents additive random noise. Our aim is to recover the 'true' image $\mathbf{x}$ from the acquired $\mathbf{y}$. The basic idea of the Bayesian approach is that the images $\mathbf{y}$ and $\mathbf{x}$ are realizations of random variables $\mathbf{Y}$ and $\mathbf{X}$. This allows us to define probability distributions on them. Based on these distributions different estimators of the 'true' image can be adopted. The first ingredient of this approach is the distribution $p(\mathbf{y}|\mathbf{x})$, called the image data model which represents the likelihood of observing $\mathbf{y}$ given the 'true' image $\mathbf{x}$. For simplicity, we assume that the random variables $Y_1, \ldots, Y_n$ are conditionally independent and that each Y_i has the same density function $f(y_i|x_i)$. In this way, the conditional density of $\mathbf{y}$ given $\mathbf{x}$ is:

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} f(y_i|x_i). \tag{1.1}$$

The model for $p(\mathbf{y}|\mathbf{x})$ depends on the imaging modality and the experimental set-up, and suitable models in our case are Gaussian, Rayleigh and Rice distributions. In Section 3.2.1 the image data model adopted in this study will be described in detail.

1.3.2 Prior model

As mentioned before, a fundamental role in the Bayesian approach is played by the prior distribution that models our 'a priori' knowledge about the true image to be estimated. This information can be quantified probabilistically by assuming that the 'true' image is a realization of a Markov Random Field with specified probability distribution $p(\mathbf{x})$. This distribution assigns high values to configurations $\mathbf{x}$ in agreement with our 'a priori' beliefs. Since $p(\mathbf{x})$ does not depend on the data, it can be defined before data acquisition and there are many different choices that can be made for it. In the simplest case we can adopt a model in which all the pixel values are assumed independent:

$$p(\mathbf{x}) = \prod_{s=1}^{n} P_s(x_s),$$

where P_s is set of probability distributions. In the most complicated case we can assume that every X_s depends on all the other pixel values X_j, $j \neq s$. These situations correspond respectively to point-wise or global knowledge about the spatial structure of $\mathbf{x}$. Because the algorithms used in practice have a computational cost related to the complexity of the adopted prior model, we need to find a good compromise between these two extreme cases. We will now see how to achieve this balance by discussing Markov Random Fields in detail.

1.3.3 Markov Random Fields

To define Markov Random Fields (MRF) we need to formalise the notion of neighbourhood. The set $\{\partial(i) : \partial(i) \subseteq S\}$ indexed by $i \in S$ is said to be a *neighbourhood system* if $i \notin \partial(i)$ and $i \in \partial(j) \Leftrightarrow j \in \partial(i)$ for all $i, j \in S$. In this case $\partial(i)$ is said to be the set of neighbouring pixels of pixel i. In this thesis we are interested in neighbourhood systems such that:

$$\partial(i) = \left\{ j : j \in S \text{ and } 1 \leq \| i - j \|^2 \leq d \right\},$$

where $d \geq 1$ and $\| i - j \|$ is the Euclidean distance between the centres of pixels i and j. Depending on the value of d, the neighbourhood system is said to be *first order* ($d = 1$) or *second order* ($d = 2$), and in this study we will use these two types of neighbourhood system. A subset c of S is defined to be *clique* if any two different pixels in c are always neighbours. We denote with C the set of all cliques. Given a neighbourhood system ∂, the prior distribution $p(\mathbf{x})$ of an image $\mathbf{x}$ is a Markov Random Field if:

$$p(x_s | x_r, r \neq s) = p(x_s | x_r, r \in \partial(s)). \tag{1.2}$$

In this way we have reduced the dependence of x_s to just the variables x_r corresponding to the neighbourhood of s.

Accordingly to the Hammersley-Clifford theorem (Winkler, 1995), we can always represent a strictly positive MRF in terms of a Gibbs distribution:

$$p(\mathbf{x}) = \frac{\exp\{-H(\mathbf{x})\}}{\sum_{\mathbf{x} \in \mathcal{X}} \exp\{-H(\mathbf{x})\}}, \tag{1.3}$$

where $\mathcal{X}$ is the set of possible images and the energy function $H(\mathbf{x}) = \sum_{c \in C} U_c(\mathbf{x}_c)$, in which U_c is a clique potential that only depends on the variables x_r with $r \in c$. In (1.3) the denominator is a normalizing constant called the partition function; often the normalizing

constant is denoted by Z. In this way, a MRF is completely determined by the energy function H. A commonly used family of prior distributions was suggested by Geman and McClure (1987). Good reviews of different models that can be adopted are given by Gilks *et al.* (1996), Besag (1989) and Winkler (1995).

1.3.4 Bayesian inference

Once the data model $p(\mathbf{y}|\mathbf{x})$ and the prior distribution $p(\mathbf{x})$ are defined Bayesian inference about the 'true' image $\mathbf{x}$ given the acquired data $\mathbf{y}$ is based on the posterior distribution $p(\mathbf{x}|\mathbf{y})$. According to Bayes theorem we have:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})}{p(\mathbf{y})}, \tag{1.4}$$

where $p(\mathbf{y}|\mathbf{x})$ is the image data model and $p(\mathbf{x})$ is the prior distribution. Since we are interested in (1.4) as a function of $\mathbf{x}$, $p(\mathbf{y})$ is constant and we can write:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})\,. \tag{1.5}$$

Based on the posterior distribution, different estimators for the 'true' image can be adopted. In particular, we now describe the Maximum A Posteriori (MAP), Minimum Mean Square (MMS) and Marginal Posterior Mode (MPM) estimators. An image configuration $\hat{\mathbf{x}}$ is called the:

- MAP estimator if it maximises $p(\mathbf{x}|\mathbf{y})$:

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

- MMS estimator if it is the mean value of the image under $p(\mathbf{x}|\mathbf{y})$:

$$\hat{\mathbf{x}} = \sum_{\mathbf{x}\in\mathcal{X}} \mathbf{x}\; p(\mathbf{x}|\mathbf{y}) = E_{p(\mathbf{x}|\mathbf{y})}\left[\mathbf{X}\right]$$

- MPM estimator if for every pixel i, $\hat{x}_i$ is a marginal posterior mode; that is, $\hat{x}_i$ maximises the marginal posterior probability of x_i given the data $\mathbf{y}$:

$$p(x_i|\mathbf{y}) \propto \sum_{x_{-i}} p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}),$$

where $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Each of these estimators minimises a particular loss functions; see Winkler (1995) for further details. In this work we adopted MAP and MMS estimator depending on the particular analysis task involved.

1.3.5 Markov chain Monte Carlo simulations

If each x_s can take any of g possible gray levels then $|\mathcal{X}| = g^n$, since $|S| = n$; for example if $n = 256 \times 256$ (this is the usual size of Magnetic Resonance images) and $g = 256$ then $|\mathcal{X}| = (10)^{157826}$ which is huge. Accordingly, for typical images it is impossible to explore the whole state space to find the MAP or MMS estimator. In addition, since the variables are not independent it is not feasible to calculate the normalizing constant of the prior distribution analytically. This is why we need Markov chain Monte Carlo (MCMC) simulations. The basic idea is that we can draw samples from a very complicated probability distribution by running an irreducible and aperiodic Markov Chain having that distribution as stationary distribution (Gilks *et al.*, 1996). Using the values that the chain takes, we can then obtain good approximations of the MAP or the MMS estimators. We now illustrate two particular Markov Chains: the Metropolis-Hastings (Metropolis *et al.*, 1953) and Simulated Annealing (Geman and Geman, 1984) algorithms.

Metropolis-Hastings algorithm

In order to estimate $E_\pi[f(\mathbf{X})]$, the expected value of a function f of $\mathbf{x}$ under a distribution of interest π such as the posterior, a Monte Carlo approach is to draw samples

$\{\mathbf{x}_t, t = 1, \ldots, N\}$ from π and then to make the approximation:

$$E_\pi [f(\mathbf{X})] \approx \frac{1}{N} \sum_{t=1}^{N} f(\mathbf{x}_t).$$

If the samples are independent, the law of large numbers ensures that this approximation can be made as accurate as desired by increasing the sample size N. However, the samples $\{\mathbf{x}_t\}$ need not necessarily be independent. They can be generated by any process which draws samples throughout the support of π in correct proportions. One way of doing this is by using a Markov chain that has π as stationary distribution. An example of such an MCMC procedure is supplied by the Metropolis-Hastings (M-H) algorithm.

The M-H algorithm is composed of two steps, the proposal and acceptance:

- **Proposal:** at each time t a candidate $\mathbf{x}'$ for the next state $\mathbf{x}^{t+1}$ is sampled from a proposal distribution $Q(\mathbf{x}|\mathbf{x}^t)$;
- **Acceptance:** the candidate $\mathbf{x}'$ is then accepted with probability:

$$\alpha\left(\mathbf{x}^t; \mathbf{x}'\right) = \min\left\{1, \frac{\pi(\mathbf{x}')\, Q(\mathbf{x}^t|\mathbf{x}')}{\pi(\mathbf{x}^t)\, Q(\mathbf{x}'|\mathbf{x}^t)}\right\}. \tag{1.6}$$

If the candidate point is accepted then $\mathbf{x}^{t+1} = \mathbf{x}'$; otherwise the chain does not move and $\mathbf{x}^{t+1} = \mathbf{x}^t$.

Since α depends only the ratio $\frac{\pi(\mathbf{x}')}{\pi(\mathbf{x}^t)}$, knowledge of any normalising constant that appears in the definition of π is not required. Different forms of proposal distribution can be adopted such as Gaussian and uniform (see Gilks *et al.*, 1996). Some iterations are required for a Markov chain to converge to equilibrium. In Section 3.2.3 we describe how we designed Q in order to improve the convergence of the algorithm. In practice all the estimations are performed using realizations that are produced after convergence has been judged to occur, with the earlier realizations of the chain being discarded. The

transient phase of the chain before convergence is known as *burn-in*. After a sufficiently long burn-in, let's say M iterations, a good approximation of the MMS estimator is given by:

$$\hat{\mathbf{x}} \approx \frac{1}{N-M} \sum_{t=M+1}^{N} \mathbf{x}_t,$$

where N is usually large. In our work, we implemented a single component M-H algorithm. In fact, instead of updating the whole image at once it is often more convenient and computationally efficient to divide $\mathbf{x}$ into components such as pixels and then to update them one by one. Hence, an iteration of the algorithm comprises n updating steps. Let x_i^t denote the value of pixel i after it has been updated during iteration t. At step i of iteration $t+1$, the M-H algorithm updates x_i as follows. A candidate x_i' is generated from the proposal distribution $Q_i\left(x_i'|x_i^t, x_{-i}^t\right)$ where $x_{-i}^t = \left\{x_1^{t+1}, \ldots, x_{i-1}^{t+1}, x_{i+1}^t, \ldots, x_n^t\right\}$, the components $1, \ldots, i-1$ having already been updated. Thus the i^{th} proposal distribution $Q_i\left(\cdot|\cdot, \cdot\right)$ generates a candidate only for the i^{th} component of $\mathbf{x}$ and may depend on the current values. The candidate is accepted with probability $\alpha\left(x_i^t, x_{-i}^t; x_i'\right)$ where:

$$\alpha\left(x_i, x_{-i}; x_i'\right) = \min\left\{1, \frac{\pi\left(x_i'|x_{-i}\right) Q_i\left(x_i|x_i', x_{-i}\right)}{\pi\left(x_i|x_{-i}\right) Q_i\left(x_i'|x_i, x_{-i}\right)}\right\},$$

where $\pi(x_i|x_{-i})$ is the full conditional distribution of x_i given x_{-i}. If x_i' is accepted, then $x_i^{t+1} = x_i'$; otherwise $x_i^{t+1} = x_i^t$. The remaining components are not changed at step i of iteration $t+1$.

Important issues in MCMC methodology are the sweep order, the burn-in length, the starting values of the chain and the number of chains to be used. These are discussed in detail in Gilks *et al.* (1996) for example. In this thesis we used a single long chain the length of which was determined by monitoring the spatial mean. The starting value of the chain was set to be an image of random numbers drawn from a uniform distribution in the range of the data. We adopted a systematic sweep order. Finally, we set the burn-in

to be the first half of the chain.

Simulated Annealing

In order to obtain a good approximation to the MAP estimator the simulated annealing algorithm is often employed. The idea is that instead of using a homogeneous Markov chain to sample from the posterior distribution $p(\mathbf{x}|\mathbf{y}) \propto \exp\{-H(\mathbf{x})\}$, we run an inhomogeneous chain by changing the target probability distribution at each iteration t to:

$$p_{\tau(t)}(\mathbf{x}|\mathbf{y}) \propto \exp\left[-\frac{H(\mathbf{x})}{\tau(t)}\right] \tag{1.7}$$

where $\tau(t) > 0$ is a parameter called the *temperature* at iteration t that tends to zero as $t \to \infty$. If $\tau(t) \to 0$ very slowly, the distribution (1.7) becomes concentrated on the MAP estimate so ensuring that the algorithm eventually yields a global minimum of $H(\mathbf{x})$. In fact, Geman and Geman (1984) established that if the temperature schedule is logarithmic, i.e. $\tau(t) = C/\log(1+t)$, the resulting Markov chain samples the MAP estimate with probability one. In practice, the theoretical value of C required to guarantee convergence is unknown and we cannot run the algorithm for an infinite time. Also, if $\tau(t)$ converges to zero too rapidly, the Markov Chain may get stuck in a local maximum. Hence the choice of a suitable temperature schedule is very important and various schedules have been studied in the literature. Some examples are the straight line, geometric, reciprocal and logarithmic schedules. These are discussed in more detail in Stander and Silverman (1994) and Hurn and Jennison (1995) for example.

Once the temperature schedule has been fixed, the simulated annealing algorithm proceeds as follows:

- at iteration t and temperature $\tau(t)$, a particular single component M-H updating step with fixed acceptance probability of 1 is performed. This is a Gibbs sampler

type step; see Geman and Geman (1984). In particular, at pixel i, x_i^t is chosen with probability:

$$p_{\tau(t)}\left(x_i^t | x_{-i}^t, y\right) = \frac{1}{Z_{\tau(t)}} \left\{ f\left(y_i | x_i^t\right) p\left(x_i^t | x_{\partial(i)}\right) \right\}^{\frac{1}{\tau(t)}},$$

where

$$Z_{\tau(t)} = \sum_{x_i^t} \left\{ f\left(y_i | x_i^t\right) p\left(x_i^t | x_{\partial(i)}\right) \right\}^{\frac{1}{\tau(t)}},$$

in which $x_{\partial(i)}$ represents the elements of x_{-i}^t at the neighbours of pixel i. Note that the calculation of $Z_{\tau(t)}$ involves the summation of only g terms and hence is computationally inexpensive.

- Once an iteration is completed, the temperature is decreased according to the chosen schedule.

1.4 Image Classification

Image classification is the process of organizing images into semantic categories. In our context, classification involves labelling every pixel of the acquired images as belonging to one of a number of possible categories or classes. A good classification method is such that pixels classified in the same category have similar values and form a connected region, while neighbouring pixels that are in different categories have dissimilar values. Classification is very important in every image-based research area in order to summarize the large amount of information collected and represent clearly the results obtained. In particular, in this thesis we study the problem of classifying pixels of the acquired images into three classes that represent three different breast tissues: normal, benign and malignant tumoral tissue. The number of classes is chosen in accordance with the opinion of radiologists. The data that we use to classify the pixels are indicated as 'class attributes' $\mathbf{d}$ and will be

described in Section 4.1. These class attributes take into account some characteristics of the temporal evolution of the signal at every pixel. Hence, based on $\mathbf{d}$, every pixel will be given a label $k \in \{1, \ldots, K\}$, where $K = 3$. For this we developed a methodology based on the K-means clustering algorithm followed by the Iterated Conditional Mode (ICM) algorithm. Our methodology does not require any training set and so we may refer to it as unsupervised classification.

1.4.1 K-means clustering algorithm

There are several variants of the K-means clustering algorithm, but most of them are based on an iterative scheme that operates over a fixed number K of clusters or classes. Each iteration involves two steps:

- for each cluster a centroid, which is the mean value of the objects in that class is found;
- each object is assigned to the class whose centroid is nearest.

The goal of the algorithm is to divide the objects into K clusters such that some metric relative to the cluster centroids is minimized. Various metrics that can be minimized include:

- maximum distance for any object to its cluster centroid;
- sum of the average distance to the centroids over all clusters;
- sum of the variance over all clusters;
- total distance between all objects and their cluster centroids.

From initial estimates of cluster centroids, every object is assigned to the cluster characterized by the closest centroid. When all objects have been assigned, the cluster

centroids are recalculated. These steps are repeated until the clusters no longer change.

The metric to minimize and the choice of a distance measure determine the shape of the final clusters. Different distance measures can be adopted such as the L_1 'Manhattan' distance or the L_2 Euclidean distance. In particular, in our context in which we consider the classification of pixels, we adopted the following distance based criterion:

$$\sum_{i=1}^{n} \left\{ \min_{k=1,\ldots,K} (x_i - \mu_k)^T \mathbf{V}^{-1} (x_i - \mu_k) \right\},$$

where x_i represents the value of pixel i, $\mu_1, \mu_2, \ldots, \mu_K$ are the estimates of cluster centroids, and $\mathbf{V}$ is the variance-covariance matrix averaged over the clusters. In this way, at each iteration the classification corresponds to the smallest total distance between all image pixels and their cluster centroids. More formally the algorithm can be described with the following four steps (Glasbey and Horgan, 1995):

1. Obtain an initial estimate of $\mu_1, \mu_2, \ldots, \mu_K$ and let $\mathbf{V}$ be the identity matrix.

2. Assign every image pixel to a cluster based on the adopted metric.

3. Re-estimate μ_k as the mean of pixels assigned to cluster k and $\mathbf{V}$ as the variance-covariance matrix averaged over clusters.

4. Repeat steps 2 and 3 until no change in the clusters is observed.

1.4.2 The Iterative Conditional Mode algorithm

The ICM algorithm was discussed by Besag (1986). We present this algorithm here in the way that it will be used later in the context of classification. From initial estimates of class means and variances a first image classification is produced. This initial classification could be provided by the K-means clustering algorithm. Next, mean pixel values are evaluated in each category and then the image is re-classified, pixel-by-pixel by optimizing

a function that takes into account both the category means and the current classification of neighbouring pixels. This process is iterated.

We now describe the ICM algorithm in detail for our classification application. Let $\mathbf{k}$ be a classified image and let k_i^t be the class of pixel i at iteration t. The ICM algorithm attempts to approximate the MAP estimator of the posterior distribution $p(\mathbf{k}|\mathbf{d})$. The idea is that at each pixel i the conditional posterior $p\left(k_i|k_{-i}, \mathbf{d}\right)$ is maximised. ICM effectively corresponds to Simulated Annealing with a temperature schedule such that $\tau(t) = 0$ (instantaneous freezing). In fact, at each iteration t of ICM, we choose k_i^t to maximise:

$$p\left(k_i^t|k_{-i}^t, \mathbf{d}\right) = \frac{1}{Z}\, p\left(d_i|k_i^t\right) p\left(k_i^t|k_{\partial(i)}\right),$$

where $p\left(d_i|k_i^t\right)$ represents the attribute model, $p\left(k_i^t|k_{\partial(i)}\right)$ is the conditional prior model and Z is the normalizing constant which needs to be estimated in our fully Bayesian approach. In Section 4.3 we present a method for estimating Z. Since $p(\mathbf{k}|\mathbf{d})$ never decreases during the algorithm, convergence is guaranteed. However, unlike the theoretical version of Simulated Annealing, convergence may be to a local minimum, so the choice of the initial point is crucial in the ICM algorithm. The whole procedure is carried out for a given number of iterations or until convergence occurs. In practice five or ten iterations are sufficient.

1.5 Deformable template models

Templates are geometrical models based on shape that are useful for high level image analysis tasks such as image segmentation, object recognition, object tracking and computer vision. Given a degraded image of a scene, the challenge is to locate and recognize one or more objects present in it. Since deterministic and random distortions affecting the data can make this task extremely difficult, it can be useful to exploit

information about the shapes of the objects that are likely to be present in the scene, whenever this is available. This information can be represented by means of geometrical models or templates. The range of application of these models is extremely wide. In particular, they play a key role in many biological and medical applications where often objects under investigation can be classified only by their shape. Early research on deformable template models concentrated mainly on rigid shape matching. In this approach the deformed shape is obtained by applying simple affine transformations (Chin and Dyer, 1986) to an initial template. The object of interest present in the scene can then be recovered using correlation based matching. In this early stage the introduction of the Hough transform (Hough, 1962) represented an important step. The classical Hough transform is a technique which can be used to isolate features of a particular shape such as lines, circles and ellipses within an image. These features must be specified in a parametric form. The generalized Hough transform (Ballard, 1981) can be employed in applications where a simple analytic description of the feature is not possible. Basically, the Hough method transforms points in the spatial feature space into a parameter space. The specified shape is then detected by finding the peaks in this space. The advantage of this method is that it is relatively insensitive to noise and occlusions[1]. The disadvantages are its limited applicability due to the high computational complexity when the number of parameters needed to describe the shape is large and the rigidness of the associated model. In fact, a rigid model is not able to recover a shape which differs from the template by transformations other than translations, rotations or scalings, the so called affine transformations. Complete surveys on different variants of the Hough Transform technique and its application can be found in Illingworth and Kitter (1988) and Leavears

[1]Geometrically, an occlusion is caused by the motion of an object that leads another object to be hidden (occluded). If the observer is moving in a static environment, occlusions corresponds to discontinuities in the perceived motion and depth. In an image sequence occlusion corresponds to a sudden lack of an expected object feature from one frame to the next.

et al. (1993).

Successive studies concentrated on templates able to deform themselves to fit the data by transformations that are more complex than affine. These models can be divided in two classes:

1. free form models;
2. parametric deformable models.

We now discuss each class in turn.

1.5.1 Free form models

Free form models, also called active contours, do not take into account the global structure of the template. The template can be any arbitrary shape as long as some regularization constraints such as smoothness and continuity are satisfied. This shape can be deformed to match salient image features such as lines and edges using potential fields produced by these features. Active contours were first proposed by Kass *et al.* (1988) and Terzopoulos *et al.* (1988). In this approach an energy minimizing contour, represented by a spline, is driven by a mixture of three forces:

- an internal spline force that enforces the smoothness;
- an image force which attracts the spline to the desired feature;
- an external constraint force.

Each force creates its own potential and the spline iteratively adjusts its position and shape until it reaches a minimum of the potential. The active contour approach has been successfully applied to edge and subject contour detection, stereo matching and image segmentation (Cohen and Cohen, 1993, Lemarie and Levine, 1993, Yuille *et al.*, 1992). A good review of free form deformable templates is given by Jain *et al.* (1996).

1.5.2 Parametric deformable models

Parametric deformable models are able to encode a specific characteristic shape and its variation. They are assumed when some 'a priori' information about the shape is available. These models can be divided into:

- analytical models, where the shape can be characterized by a collection of parameterized curves;
- prototype models, where a prototype is assumed and it is deformed by a parametric mapping.

In the first case the template is represented by a set of curves that are uniquely described by some parameters. In this way, the specific analytical form incorporates the prior knowledge about the shape of the objects under analysis. The parametric shape of the template can be changed by varying the parameter values. A potential energy field is defined based on the image features and the current template and the shape evolves to minimize this energy by updating the parameters. In this framework Grenander *et al.* (1993) formulated a global pattern-theoretic model of shape. This consists of:

- a space of generators such as pixels, polygons, vertices and edges which are the basic building blocks of the structure;
- a connector graph which describes the interaction between generators;
- bonding relations which apply geometric constraints so that the resulting configurations are meaningful;
- a transformation group which produces new structures by mapping one generator into another.

In general, this approach can be represented as the composition of a model template and a parametric statistical mapping. The former describes the overall architecture of the shape and the latter governs the random variations in the building blocks of the shape. These factors together should be able to control the desired global and local geometry of the shape class.

Some other examples of parametric templates are given by elliptic Fourier descriptors, moment invariants based models and eigenshapes models (see Hurn, 1998, Baumberg and Hogg, 1995, Haddadnia *et al.*, 2001, for example).

In general, the applicability of parametric deformable models is somewhat limited because the shapes under investigation must be well defined so that they can be represented by a set of curves characterized by a small number of parameters. Depending on the application this is not always achievable.

These problems can be overcome with a prototype model in which a prototype template is deformed by a parametric mapping. An objective function based on the image features and the edge of the template is defined and the mapping parameters are then iteratively updated in order to minimize this function. The prototype template can be either specified by means of expert advice, for example a sketch drawn from a physician, or obtained from training samples. The chosen parametric statistical mapping reflects the particular deformations allowed in the application domain. Some typical deformations are represented by rubber sheet deformations based on a trigonometric basis, thin plate splines, Kriging (Dryden, 1998) or Wavelets (Abramovich *et al.*, 2000).

We should emphasize that recovering the shape of one or more objects from noisy images by using prototype models consists of more than removing the noise from the data. It provides a transformation of the template which gives a structural understanding of the picture. In other words, the transformation allows us to identify characteristic

features of the true object in the image by making comparisons with those of the template. For example, this approach could lead us to determine objectively landmarks and their geometric parameters in the object. In medical applications we may be able to identify or even classify pathologies.

In the framework of deformable template models, the Bayesian approach plays an important role, as it always does when prior knowledge of a process is available that needs to be combined with acquired data to make statistical inference about the parameters of the process. The prior distribution typically represents knowledge about the objects in a particular scene and the image data model represents the joint probability distribution of the acquired data conditional on the objects present in the scene. These two ingredients are combined to obtain the posterior distribution that represents the probability of observing a particular deformed template given the acquired data. Inference about the parameters of the mapping are then obtained from the posterior distribution. The MAP and MMS estimators are often adopted to estimate the optimal deformation of the template.

Chapter 2

The Data

In this chapter we describe the data analysed in this thesis. In Section 2.1 we introduce briefly the physical principles of Nuclear Magnetic Resonance and Magnetic Resonance Imaging. In Section 2.2 we describe the experimental set-up and the actual data motivating the methodology that we developed.

2.1 The physics of Magnetic Resonance Images

Magnetic Resonance Imaging (MRI) is a non-invasive technique that can produce high resolution images of slices of living tissues. This technique is based on the principle of Nuclear Magnetic Resonance (NMR) which describes the interaction of radio waves with atomic nuclei that have particular values of a property called spin. Such nuclei are called paramagnetic[1]. This principle allows us to probe objects with radio waves. In particular, we can obtain physical information by collecting the signal that is emitted when the spins that have been excited by radio waves release energy upon 'relaxing' back to the equilibrium state. In fact, nuclear energy levels are quantized with the consequence that to induce a transition from one energy state to another, the exact amount of energy corresponding to the difference between the two states must be absorbed by the nucleus.

[1]Paramagnetic nuclei have an odd number of electrons in the outer electron shells.

In NMR the nucleus absorbs this energy from radio waves. Since the radio wave energy is proportional to its frequency, we can cause transitions between two states of a particular nuclear species by setting the radio wave frequency to the corresponding energy gap.

2.1.1 Nuclear Magnetic Resonance principles

The most widely used isotope in MRI is the Hydrogen (H1) atom since its high abundance in living tissue yields high MR signals. The nucleus of this atom can exist in two different energy levels that coincide in the absence of a static magnetic field. However, if a static magnetic field B_0 is switched on, these energy levels split and are characterized by an energy gap. If we now apply a radio frequency pulse corresponding exactly to this energy gap, we induce transitions between the two energy levels. Once the radio frequency is switched off the nuclei lose coherence and relax back to their original state emitting energy. The emitted energy gives rise to the acquired NMR signal. This process can be described classically in the following way. When we place a sample of paramagnetic nuclei in the magnetic field, the magnetic moments of each nucleus sum together to produce the net magnetization M. When B_0 is switched on, M revolves or precesses around it at a particular frequency, known as the precession frequency that depends on the applied magnetic field. We say that the nuclei *resonate* with this frequency and hence speak of Nuclear Magnetic *Resonance*. The radio frequency pulse can be represented as a different magnetic field B_1 perpendicular to B_0. Hence when the radio frequency pulse is on, M tends to precess also around B_1 with the result that the magnetization axis now forms an angle α, called the flip angle, with B_0. The value of α is proportional to the time of application of this radio frequency pulse. When the radio frequency pulse B_1 is switched off, M gradually relaxes to realign with B_0 and energy is emitted. At this stage the NMR signal is collected. This signal depends on the projected components of the

magnetization on the axis of the plane perpendicular to B_0 and on the radio frequency pulse. The technique that we have described is used to detect nuclei without any spatial discrimination; that is the acquired NMR signal depends on all nuclei whose energy levels are separated by a gap corresponding to the frequency of the radio wave applied. This is the basis of NMR spectroscopy. We now describe how we can extend this technique to NMR imaging, in which the acquired signal depends on the spatial coordinates of the nuclei.

2.1.2 Magnetic Resonance Imaging

In order to have a MR image we need to discriminate between signals that depend on different magnetization coordinates (x, y, z). Let us assume that the static magnetic field B_0 is oriented along the z axis so that the MR signals will depend on the (x, y) components of M. The idea behind MRI is to generate signals in which different spatial locations produce contributions at different frequencies and phases. This is achieved through the use of linear variations in the strength of the magnetic field B_0 in the x, y and z directions with the result that the precession frequency becomes a function of space. This spatial information is encoded into the signal using gradients by means of three separate processes: slice selection, frequency encoding and phase encoding. These are presented graphically in Fig. 2.1. We will now describe each step briefly.

Imaging a slice of the object under investigation requires a method to excite nuclei only within that slice. To do this, in the slice selection step, a linear gradient along the z axis is applied to B_0 so that the precession frequency varies along z. In this way, if we apply a radio frequency pulse containing just a small range of frequencies, only nuclei in the slice corresponding to these frequencies will be excited.

Further spatial information is embedded into the signal during the frequency encoding

step by applying a linear gradient along the x axis as the signal is acquired. This causes nuclei with different x coordinates to precess at different frequencies, and thus to produce signals that have different frequencies when collected.

Finally, in the phase encoding step, spatial information in the y direction is embedded in the signal by reading it on K separate occasions under slightly different conditions. On the k^{th} occasion a linear gradient in the y direction is employed in between the application of the z and x gradients at time t_k. This causes the frequency of the spins to change temporarily according to the y coordinate and results in a corresponding change of the phase in the y direction. The time t_k varies on each application so that a different amount of phase is encoded at each y coordinate. This step completes the coding of the spatial information in the acquired signal.

A complete overview of MRI techniques is given in Stuart and Young (1988). In MRI different choices of radio frequency pulses and gradients can be adopted depending on the particular application. The particular paradigm adopted for acquiring the image is usually called the MRI sequence. In dynamic MRI studies a Fast Low Angle SHot (FLASH) is usually employed (Matthaei *et al.*, 1992). In this sequence the flip angle α is small in order to reduce the acquisition time. Unfortunately this leads to noisy image sequences.

2.2 The analysed dataset

The data analyzed in this project were acquired at the Radiology department of the 'Istituto Regina Elena' in Rome and consist of a dMRI sequence. In this thesis, we consider a typical dMRI sequence of 20 images of 256 × 256 pixels. The spatial and temporal resolution are approximately 1 mm^2 and 15 seconds respectively and the slice thickness is 5 mm. The images are acquired by a Siemens Magnetom MR tomograph operating at 1.5 Tesla. In Fig. 2.2 we show the acquired image sequence in lexicographical

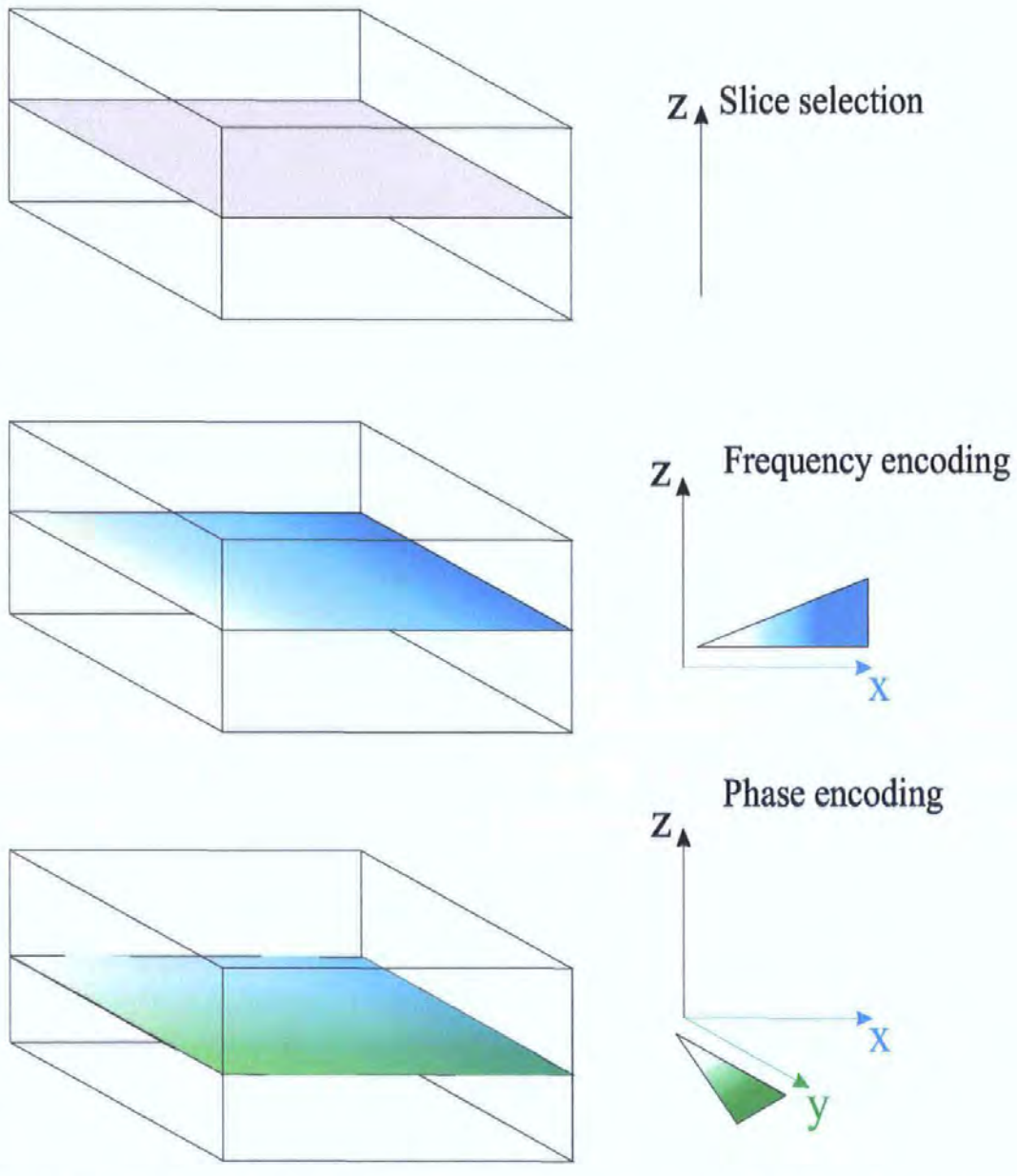


Figure 2.1: *MRI spatial encoding. Slice selection: a gradient is applied along the z axis to the static magnetic field B_0 so that for a particular value of the frequency only spins lying on a plane (gray) will be selected. Frequency encoding: a second gradient (blue) is applied along the x axis so that spin frequencies vary in the x direction. Phase encoding: a gradient is applied on the y axis (green) inducing a change of phase in this direction.*

order. The first image of the sequence is acquired immediately after the contrast agent injection. The image intensity depends on the gadolinium concentration and in Fig. 2.3 we can observe the signal increase over time between the first two images of the series. This increase is significant within the Region Of Interest (ROI) indicated by the box. As will be described in Section 3.1, we chose to apply our Bayesian methodology within this ROI because a preliminary step of our analysis, presented in Section 3.1, will reveal significant temporal signal variations inside this region. In the reported dMRI sequence we note a fast increase of intensities within the region corresponding to the selected ROI in the first three images of the sequence. The following images are characterized by relatively small intensity variations. This behaviour reflects the typical uptake pattern of the contrast agent described in Section 1.2 and shown in Fig. 1.1. This consists of a very fast signal increase followed by either a plateau or a low signal decrease. This behaviour can be observed in Fig. 2.4(a) where the temporal pattern for a given pixel within the ROI is shown. In particular, from this figure the steep uptake observed during the first three images and the slow decrease during those that follow are evident.

Two main sources of distortions affect the acquired data set: deterministic distortions due to the patient movements during image acquisition and random distortions due to the experimental set-up. The method used to reduce the deterministic distortions due to patient movements will not be described in this work. In the following, we will show the data after deterministic distortions have been minimized.

In Fig. 2.4(a) the random distortions affecting the data are evident. Since the classification task is based on the temporal evolution of the signal, it is important to minimize the effect of these random fluctuations. In Fig. 2.4(b) we report the last image of the sequence for the ROI. Again we note that this image is clearly affected by random distortions that potentially reduce the diagnostic capability of dMRI.

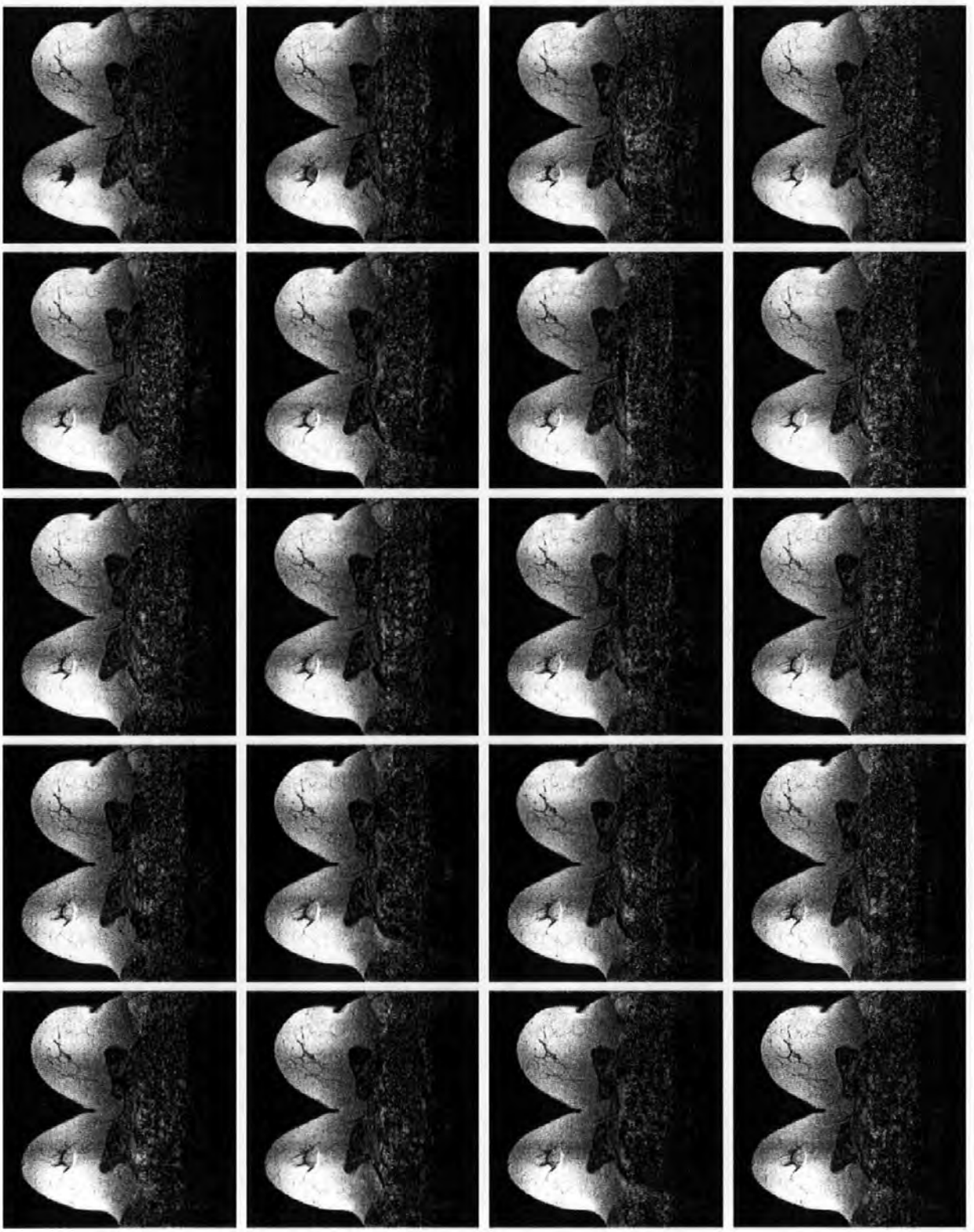
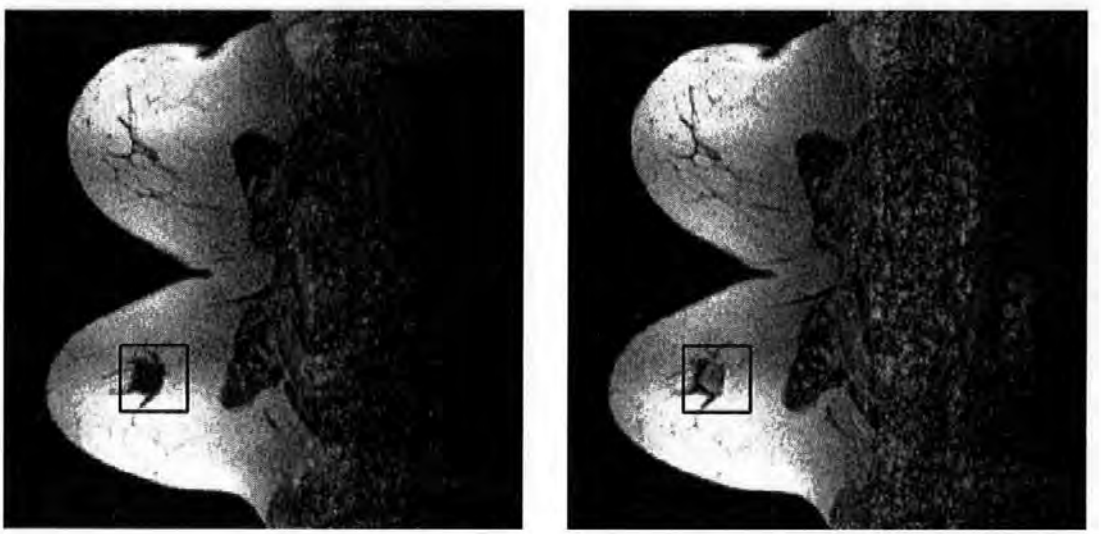
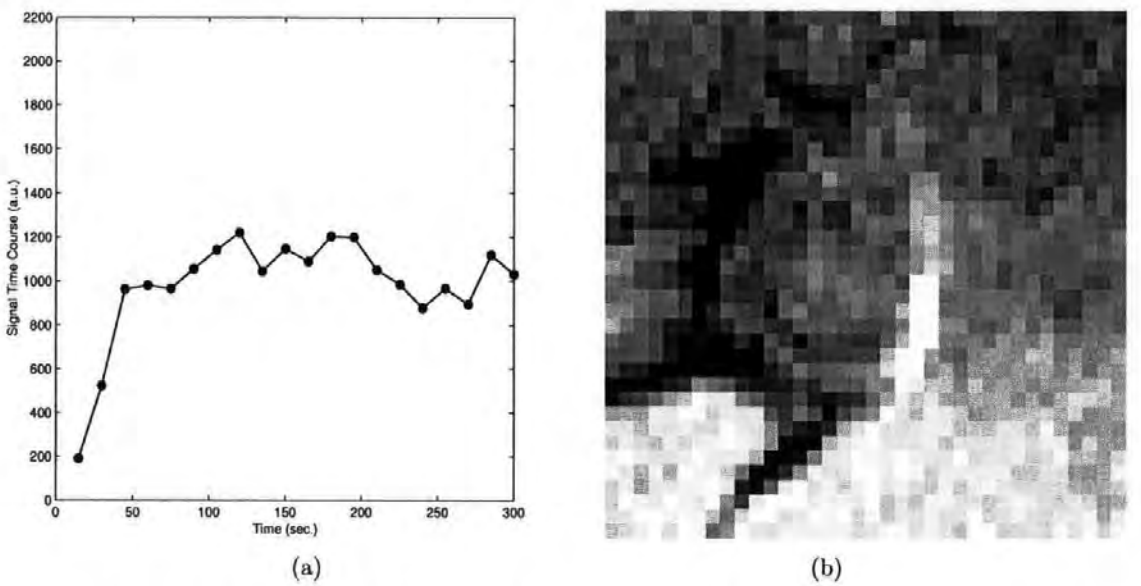


Figure 2.2: *The acquired dMRI sequence. The sequence is shown in lexicographical order, i.e. the top left frame is the first image of the sequence acquired immediately after the contrast agent injection. The bottom right frame is the last image of the sequence.*



(a) (b)

Figure 2.3: *Two images from the acquired dMRI sequence: (a) first image of the sequence acquired immediately after the contrast agent injection; (b) second image of the sequence acquired 15 seconds after the contrast agent injection. The ROI is indicated by a box.*



(a) (b)

Figure 2.4: *(a) A typical time pattern for a pixel within the ROI. Random distortions make the classification of ROI pixels from their temporal evolution extremely difficult. (b) The selected ROI in the last image of the sequence. The random distortions reduce the capability of this technique to investigate the internal structure of the lesion.*

Chapter 3

Bayesian Image Restoration

In this chapter we present a parametric and a non-parametric Bayesian approach for the image restoration task, together with a novel solution to the problem of estimating the model hyper-parameters and a modified version of the Metropolis Hastings algorithm.

We will begin by introducing some further notation: let $\mathbf{y}_i = \{y_i(1), \ldots, y_i(T)\}$ represent the observed temporal intensity profile at pixel i, where $i = 1, \ldots, n$, and let $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$ be all the observed data, that is a sequence of T images. Our values of n and T, as mentioned in Section 2.2 are 256×256 and 20 respectively. Similarly, let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ be the true, but unobserved image sequence to be estimated. In order to reduce considerably the computational burden, our Bayesian analysis will not be performed on the entire images of the sequence but on a selected ROI from them consisting of a few thousand pixels.

This chapter is structured as follows. In Section 3.1 we describe how the ROI is selected and how we use an hypothesis test based approach to locate the tumoral region. Then the non-parametric approach is introduced in Section 3.2, while in Section 3.3 the parametric method is presented. Finally, in Section 3.4 the results of the tumoral region location procedure are given and the image restorations obtained with both methods are shown and compared.

3.1 Region Of Interest selection and tumoral region location

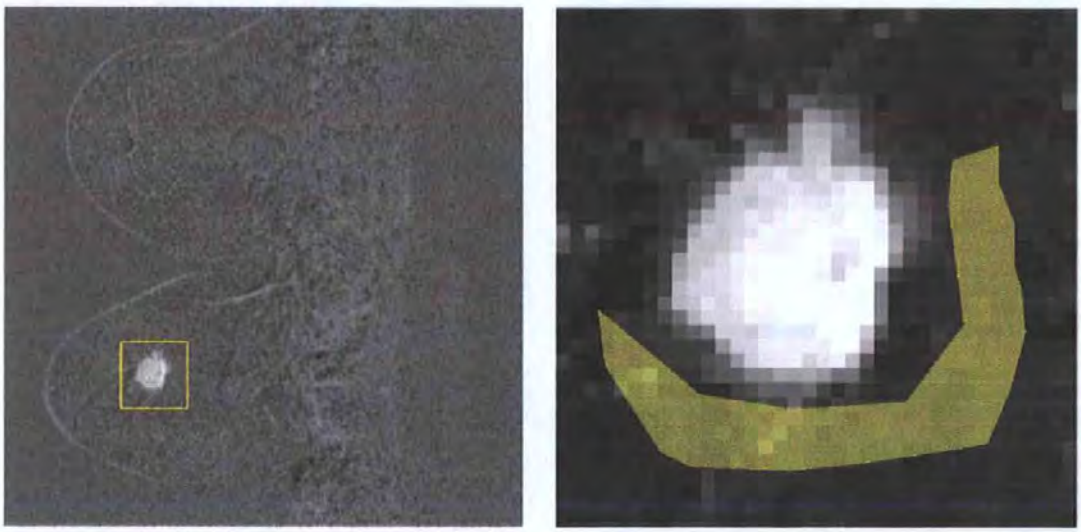
In this thesis we assume that there is only one lesion in the breast. If this is not the case both the image restoration and classification procedures have to be repeated for the additional lesions that may be present.

The very first step of our dMRI analysis is to identify a ROI within the breast where a 'significant' increase of the contrast agent concentration is observed. In order to find this region from the acquired image sequence $\mathbf{y}$, we start by calculating the 'mean difference image' $\mathbf{m}$, the i^{th} pixel of which is given by:

$$m_i = \frac{1}{T-1} \sum_{t=2}^{T} \{y_i(t) - y_i(1)\}, \tag{3.1}$$

where $y_i(t)$ is the image acquired at time t and $t = 1, \ldots, T$. By subtracting $\mathbf{y}(1)$ from every image of the sequence we reduce considerably the dependency of $\mathbf{m}$ on the anatomical features of breast. In fact, since $\mathbf{y}(1)$ is acquired immediately after the Gd-DTPA injection, the contrast agent concentration in it is still so low that the signal at every pixel will depend only on the MR parameters of the tissue in that location before the contrast agent injection. The image $\mathbf{m}$ is an important parameter in this kind of analysis because nowadays physicians base their diagnosis on it. In Fig. 3.1(a) we report $\mathbf{m}$. We note the presence of a bright region where the high signal value is due to the increased concentration of the gadolinium salt. In this patient we can clearly identify a lesion. In fact, physicians identify the tumoral lesion location by extracting the contour of the bright region within $\mathbf{m}$; we present the results of this procedure in Section 4.4.

Based on $\mathbf{m}$, a ROI containing the identified bright region can be drawn by the radiologist. In Fig. 3.1(a) we show a typical ROI (yellow box) superimposed on $\mathbf{m}$. In the following, our analysis will be focused on the pixels in this ROI. In this way, the number of



(a) (b)

Figure 3.1: *(a) The 'mean difference image'* $\mathbf{m}$ *and the selected ROI (yellow). Here we note how dependency on the anatomical features of the breast has been considerably reduced and contrast agent uptake highlighted. The bright area inside the ROI corresponds to a tumoral region. (b) The image* $\mathbf{m}$ *in the selected ROI with the reference region* $\mathcal{R}$ *(yellow) superimposed. The region* $\mathcal{R}$ *comprises healthy tissue surrounding the lesion.*

pixels to be analyzed decreases from approximately one million to a few thousand (usually the size of the ROI is about 50×50 pixels). This obviously makes our procedures faster.

The next step is to locate more precisely the lesion within the ROI. To do this we implemented an hypothesis testing procedure. First, a reference region $\mathcal{R}$ within the ROI is selected by the user. This represents a region where the user is confident that there is no significant glucose uptake. In Fig. 3.1(b) we show a typical $\mathcal{R}$ within the ROI of Fig. 3.1(a). This region should comprise healthy tissue surrounding the tumour. In order to reduce variability, $\mathcal{R}$ is selected in the same breast where the lesion has been identified. Next, the distribution of pixel values within $\mathcal{R}$ is estimated. Because of the shape of this distribution, shown later in Fig. 3.5, we decided to fit this empirical distribution with a Gaussian model the mean and the standard deviation of which have to be estimated. As these parameters are well estimated, it is sufficient for our purposes to assume that they are known; we will denote them by μ_R and σ_R. In this way we are now able to perform an hypothesis test over the pixels within the ROI. In particular the hypotheses at pixel i are:

$$\begin{aligned} &\mathrm{H}_0\text{: pixel } i \text{ is non-tumoral} \\ &\mathrm{H}_1\text{: pixel } i \text{ is tumoral,} \end{aligned}$$

so that the test statistic will be:

$$\begin{aligned} \mathrm{H}_0\text{:} \quad & \frac{M_i - \mu_R}{\sigma_R} \leq z_\alpha \\ \mathrm{H}_1\text{:} \quad & \frac{M_i - \mu_R}{\sigma_R} > z_\alpha \quad , \end{aligned}$$

where z_α is such that $\mathrm{P}\left(Z \leq z_\alpha\right) = 1 - \alpha$, in which $Z \sim N(0,1)$. The probability of Type I error or significance level is approximately α. The basic idea behind this test is that tumoral pixels are characterized by a distribution that is different from $N(\mu_R, \sigma_R^2)$. In

particular, since we expect the mean value of tumoral pixels over time to be considerably greater than μ_R, we used a one-tail test. Pixels for which H_0 is rejected in favour of H_1 will be classified as tumoral. In the following, we will refer to the set of tumoral pixels as the tumoral region which we will denote $\mathcal{I}$. We will discuss results of this hypothesis test procedure in Section 3.4.

3.2 A non-parametric approach for image restoration

The Bayesian estimation of the true image intensities is performed at each pixel within the ROI shown for example in Fig. 2.4(b) for the last image of the sequence. This is done without adopting a parametric temporal model for the image intensity evolution. Following the Bayesian paradigm, we now introduce the image data model, the prior model and the adopted estimator based on the posterior distribution.

3.2.1 Image data model

Since we assume that deterministic degradations due to patient movements during the image acquisition have already been reduced, the acquired image sequence $\mathbf{y}$ will be related to $\mathbf{x}$ by:

$$y_i(t) = x_i(t) + \epsilon_i(t), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T \tag{3.2}$$

where the errors $\epsilon_i(t)$ are assumed to be independently distributed. The time unit is the time interval between two consecutive images, although in the figures we report the corresponding seconds. For an MR image acquired in absolute value the distribution of $\epsilon_i(t)$ in the background, i.e. a region of very low signal intensity outside the imaged part of the body, is known to be a Rayleigh with variance s^2 (Henkelman, 1985 and Sijbers *et al.*, 1999). In regions of high signal, the distribution of $\epsilon_i(t)$ can be well approximated

by a Gaussian with variance $\sigma^2 = \frac{4-\pi}{2}s^2$ (de Pasquale *et al.*, 2000). The value of s^2 can estimated very accurately in the background of the image. In this way, from the estimate of s^2 we can obtain σ^2 so that in the following this will be considered a known parameter. We experimented with both distributions and, since no significant differences were obtained, we adopted the Gaussian distribution for computational simplicity.

3.2.2 Prior model

Our 'a priori' distribution for the true images models both the continuity of the temporal evolution at each spatial location and the presence of homogenous spatial regions, separated by discontinuity lines, within every image of the sequence. To achieve this, we relate stochastically in a separate way the differences between image intensities at contiguous times at each pixel and the differences between neighbouring pixels in space at each time. We do this by means of the following factorized Markov Random Field model:

$$\begin{aligned} \mathrm{P}(\mathbf{x}) \quad \propto \quad & \prod_{t=1}^{T} \exp\left[-\beta_s \sum_{<ij>} V_s\left\{x_i(t) - x_j(t)\right\}\right] \times \\ & \prod_{i=1}^{n} \exp\left[-\beta_t \sum_{<t't''>} V_t\left\{x_i(t') - x_i(t'')\right\}\right], \end{aligned} \tag{3.3}$$

where V_l is the prior potential in space or time, $l \in \{s,t\}$, $< ij >$ indicates second order neighbour pixels in space, $< t't'' >$ indicates first order neighbour pixels in time, β_l is the smoothing hyper-parameter in space or time. The prior distribution (3.3) is a pair-wise interaction model characterized by the prior potentials V_s and V_t. In particular, we take:

$$V_l(z) = -\log\left\{p_l(z)\right\}, \tag{3.4}$$

where p_l is the distribution of grey level differences in $\mathbf{x}$ (Sebastiani and Godtliebsen, 1997). We model $p_l(z)$ as:

$$p_l(z) \propto \left\{1 + \left(\frac{z}{\delta_l}\right)^2\right\}^{-1}, \tag{3.5}$$

where δ_l, $l \in \{s, t\}$, are two further hyper-parameters to be estimated. The meaning of δ_l will be discussed in Section 3.2.4. With this choice we have:

$$V_l(z) = \log\left\{1 + \left(\frac{z}{\delta_l}\right)^2\right\}. \tag{3.6}$$

Model (3.6) penalizes differences depending on their amplitudes compared to the parameter δ_l. In order to illustrate this effect in Fig. 3.2 we report the behaviour of $V_s(\Delta y_s)$ corresponding to two different values of δ_s ($\delta_s = 10$ and $\delta_s = 30$, since the range that we adopted later for the optimization of this parameter is $[10, 100]$. For a given value of δ_s the differences between neighbouring values are penalized depending on their amplitudes. In particular, the larger the absolute value of Δy_s the higher the penalty V_s. As we can see from Fig. 3.2, increasing the value of δ_s leads to a lower penalty for any given differences Δy_s. Therefore, informally increasing δ_s increases the range of differences that are allowed under the prior, thus we can say that δ_s controls the amplitude of the discontinuities to be preserved. Other choices for V_l have been proposed in literature with similar behaviour (Künsch, 1994). The key feature of our approach is to define V_l through p_l by (3.4), as this allows us to propose a successful procedure for estimating δ_l. Among the different models we tried for $p_l(z)$, the one in (3.5) leads to the most reliable hyper-parameter estimation results.

3.2.3 Estimation

As we mentioned in Section 1.3.4, in the Bayesian approach different estimators for the true images can be adopted. At this stage we choose the MMS estimator because it

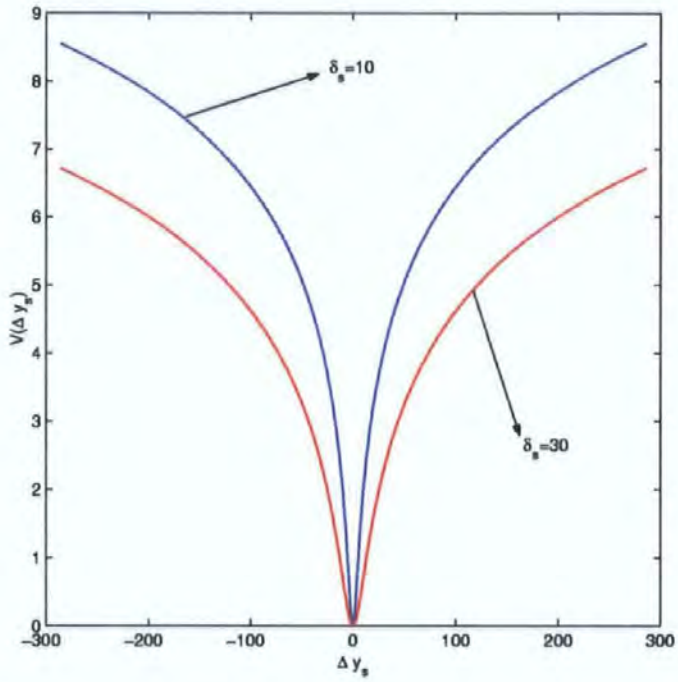


Figure 3.2: *The behaviour of the prior potential $V_s(\Delta y_s)$ corresponding to different values of δ_s: $\delta_s = 10$ (blue line) and $\delta_s = 30$ (red line). The value of δ_s controls the amplitude of the discontinuities to be preserved.*

can be obtained without solving an optimization problem. In order to obtain a good approximation to the MMS estimator we used a specially tailored single component M-H algorithm with a proposal function based on the data distribution. In particular, Q is taken to be a normal distribution with expected value $y_i(t)$ and variance equal to the estimated value of σ^2. In order to compare this proposal function with other choices, namely a uniform independence proposal and a random walk with Gaussian proposal optimized with respect to its variance, we performed a simulation study. We assumed as true image sequence real MR images characterized by a low level of noise. Then we generated a noisy dataset by adding Gaussian noise to this image sequence. Finally, we restored this image using the different M-H algorithms that we considered. To estimate the speed of convergence of these algorithms we plot the Mean Square Error (MSE) between the true image and the current approximation of the posterior mean as a function of MCMC iteration number. In Fig. 3.3 we show this comparison for the three proposal distributions that we considered. As we can see from this figure, our algorithm reaches convergence much earlier than the other two. We observed the same behaviour for different levels of noise and for other types of original images. We also remark that an update of the proposed algorithm is computationally less expensive than an update of the other algorithms. This is because the acceptance probability of the proposed algorithm has a contribution from only the prior, whereas the acceptance probability of the other algorithms has a contribution from both the prior and the likelihood.

The starting point of the algorithm is an image of random numbers drawn from a uniform distribution in the range of $\mathbf{y}$ and the pixel sweeping order is systematic. We approximate $\hat{\mathbf{x}}$ as the mean of the last half of the Markov chain sequence of images, the length of which is determined by monitoring the spatial mean.

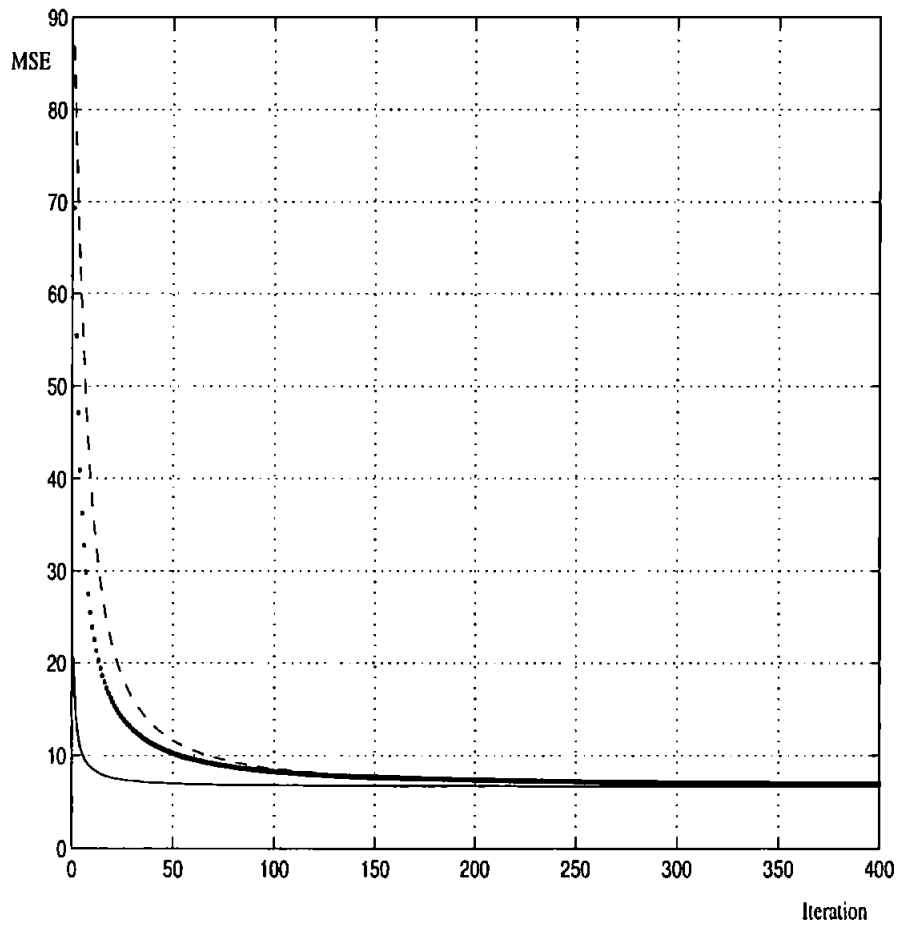


Figure 3.3: *MSE comparisons between Metropolis algorithms with different proposals: optimal Gaussian random walk (dashed line), uniform independence proposal (dotted line) and Gaussian independence proposal with measured pixel image intensity as proposal mean and the estimated value of σ^2 as proposal variance (continuous line). Each iteration on the horizontal axis corresponds to a full update of the ROI. The vertical axis represents the MSE between the true image and the current estimate of the posterior mean.*

3.2.4 Hyper-parameter estimation

One of the advantages of our Bayesian methodology is that the hyper-parameters are estimated automatically. We now describe this important step. All the hyper-parameters $(\beta_s, \beta_t, \delta_s, \delta_t)$ play an important role in our procedure. The parameters β_s and β_t represent the weights of the spatial and temporal prior potentials with respect to the image data model potential. Thus, inappropriate values of these parameters can lead to either over-smoothed or very noisy restorations. The parameters δ_s and δ_t are also very important because they control the amplitude of the discontinuities that will be preserved during the restoration. Since these discontinuities identify the location of the tumoral lesion and characterize its structure, it is crucial to estimate the δ_l accurately.

We begin by estimating δ_s and δ_t. Our estimation procedure is based on minimizing the difference between $\hat{\mathrm{p}}_l(\Delta y_l)$, the empirical distribution of observed image differences, and $\mathrm{p}_l(\Delta y_l)$, the associated theoretical distribution. The quantity Δy_l represents all the spatial or temporal intensity differences in neighbouring pixels of the observed sequence $\mathbf{y}$ (Sebastiani and Godtliebsen, 1997). From (3.2) we have that $\Delta y_l = \Delta x_l + \Delta \epsilon_l$, from which it follows that:

$$\mathrm{p}_l(\Delta y_l) = \mathrm{p}_l(\Delta x_l) \otimes \mathrm{p}_l(\Delta \epsilon_l) \tag{3.7}$$

where $\otimes$ indicates the convolution integral. Since we assumed that ϵ_l belongs to a Gaussian distribution, it follows that $\mathrm{p}_l(\Delta \epsilon_l)$ is a $N(0, 2\sigma^2)$. Based on $\mathrm{p}_l(\Delta x_l)$ from (3.5) and $\mathrm{p}_l(\Delta \epsilon_l)$, $\mathrm{p}_l(\Delta y_l)$ can be computed by performing the convolution (3.7) numerically. Finally, to estimate δ_l we minimize the sum of the absolute values of the differences between $\hat{\mathrm{p}}_l(\Delta y_l)$ and $\mathrm{p}_l(\Delta y_l)$ over parameter δ_l.

We adopted a different criterion to obtain β_s and β_t. These hyper-parameters are estimated by minimizing the discrepancy between a theoretical χ^2_T distribution and the

empirical distribution of the values in the set:

$$\Sigma_{\mathcal{I}} = \left\{ \sum_{t=1}^{T} \frac{\{y_i(t) - \hat{x}_i(t; \beta_s, \beta_t)\}^2}{\sigma^2}, \quad i \in \mathcal{I} \right\}.$$

Here $\hat{x}_i(t; \beta_s, \beta_t)$ represents the estimated image at pixels in $\mathcal{I}$ for a given choice of β_s and β_t. The rationale behind this criterion is that the distance between the empirical distribution of the values in $\Sigma_{\mathcal{I}}$ and the theoretical χ^2_T distribution decreases as $\hat{x}_i(t; \beta_s, \beta_t)$ becomes closer to true value $x_i(t)$. In order to reduce the computational time, we adopted the following scheme to minimize the discrepancy under consideration:

1. an initial point (β_s^0, β_t^0) and an incremental pass $(\Delta\beta_s, \Delta\beta_t)$ are fixed using our experience of suitable values of these parameters for the type of images under investigation.

2. The values $(\beta_s^0 + \ell\Delta\beta_s)$ are calculated where $\ell \in \{-1, 0, 1\}$. The value of the three β_s points that minimizes the discrepancy between χ^2_T and the distribution of $\Sigma_{\mathcal{I}}$ is chosen as β_s^1.

3. Analogously, $(\beta_t^0 \pm \Delta\beta_t)$ are calculated and the value of the three β_t points minimizing the discrepancy between χ^2_T and the distribution of $\Sigma_{\mathcal{I}}$ is chosen as β_t^1.

4. Steps 1 and 2 are repeated until one of the minimizing points $\left(\beta_s^k, \beta_t^k\right)$ is chosen for the second time.

We compared the parameter estimates produced by this scheme with those yielded by an exhaustive search over a fine grid of parameter values. No substantial differences emerged.

3.3 A parametric approach for image restoration

As an alternative to the non-parametric approach we now investigate a different restoration method in which a parametric temporal model for the true image intensity profile at each pixel is adopted. The parameters of this model are estimated by the Bayesian approach and some of them are used as attributes in the subsequent image classification step described in Chapter 4. We now discuss the image data model, the prior model and the adopted estimator for these model parameters.

3.3.1 Image data model

As we stated in Section 1.2, the injection of the contrast agent gd-DTPA gives rise to a significant increase in the MR signal received from regions characterized by high blood vascularization. A mathematical model describing the MR signal variation over time is very important for performing quantitative analyses of the contrast agent concentration. Unfortunately, the complexity of the physics behind dynamic MR images prohibits the use of a completely detailed model. Among the few suitable models described in the literature, a pharmacokinetic model, initially proposed by Tofts and Kermode (1991) for MRI measurements on the blood-brain barrier, has proved very important. This model has since been developed for use with dMRI sequences; see Brix *et al.*, (1991). Their model is compartmental, with the blood plasma being approximated as one compartment and the extracellular space of the tissue being a peripheral compartment. Hayton *et al.* (1996) solve the resulting pharmacokinetic equations using the Laplace Transform in such a way that different injection modalities can be taken into account. The parametric model that we present has a very similar behaviour to the solution presented by Hayton *et al.* (1996) when an instantaneous injection of the contrast agent is considered. In particular, we assume that $x_i(t)$ takes the functional form $x_i(t) = f_{\boldsymbol{\theta}_i}(t)$, where the temporal model

$f_{\boldsymbol{\theta}_i}(t)$ is given by:

$$f_{\boldsymbol{\theta}_i}(t) = \begin{cases} I_i + \dfrac{(M_i - I_i)}{1 - \exp\{-\Delta(p_i - 1)/\tau_i\}} \left[1 - \exp\{-\Delta(t-1)/\tau_i\}\right] & 1 \leq t \leq p_i \\ M_i - \dfrac{M_i - F_i}{T - p_i}(t - p_i) & p_i \leq t \leq T \end{cases}$$

in which $\boldsymbol{\theta}_i = (I_i, M_i, F_i, p_i, \tau_i)$ represents the parameter vector for pixel i, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$ is the sequence of parameter vectors for the n pixels, and $\Delta = 15$ seconds is the temporal interval between two subsequent images so that the units of τ_i are seconds. These parameters are illustrated in Fig. 3.4.

Combining the parametric model with the Gaussian noise distribution, we obtain the following image data model:

$$\mathrm{P}(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{nT/2}} \prod_{i=1}^{n} \prod_{t=1}^{T} \exp\left[-\frac{\left\{y_i(t) - f_{\boldsymbol{\theta}_i}(t)\right\}^2}{2\sigma^2}\right].$$

3.3.2 Prior model

The 'a priori' model adopted is similar to the one adopted for the non-parametric approach. In fact, for each parameter we take into account spatial continuity and the presence of different structures in the image. Among the different types of dependent models with which we experimented, the one that performs best is

$$\mathrm{P}(\boldsymbol{\theta}) = \mathrm{P}_{I,M,F}(I, M, F) \times \mathrm{P}_{p,\tau}(p, \tau)$$

where:

$$\begin{aligned} \mathrm{P}_{I,M,F}(I, M, F) &\propto \exp\left[-\beta_1 \sum_{<ij>} \log\left\{1 + \left(\frac{I_i - I_j}{\delta_I}\right)^2 + \left(\frac{M_i - M_j}{\delta_M}\right)^2 + \left(\frac{F_i - F_j}{\delta_F}\right)^2\right\}\right] \\ \mathrm{P}_{p,\tau}(p, \tau) &\propto \exp\left[-\beta_2 \sum_{<ij>} \log\left\{1 + \left(\frac{p_i - p_j}{\delta_p}\right)^2 + \left(\frac{\tau_i - \tau_j}{\delta_\tau}\right)^2\right\}\right] \end{aligned} \tag{3.8}$$

where $< ij >$ indicates second order neighbours in space and β_1, β_2, δ_I, δ_M,δ_F, δ_p and δ_τ are the hyper-parameters.

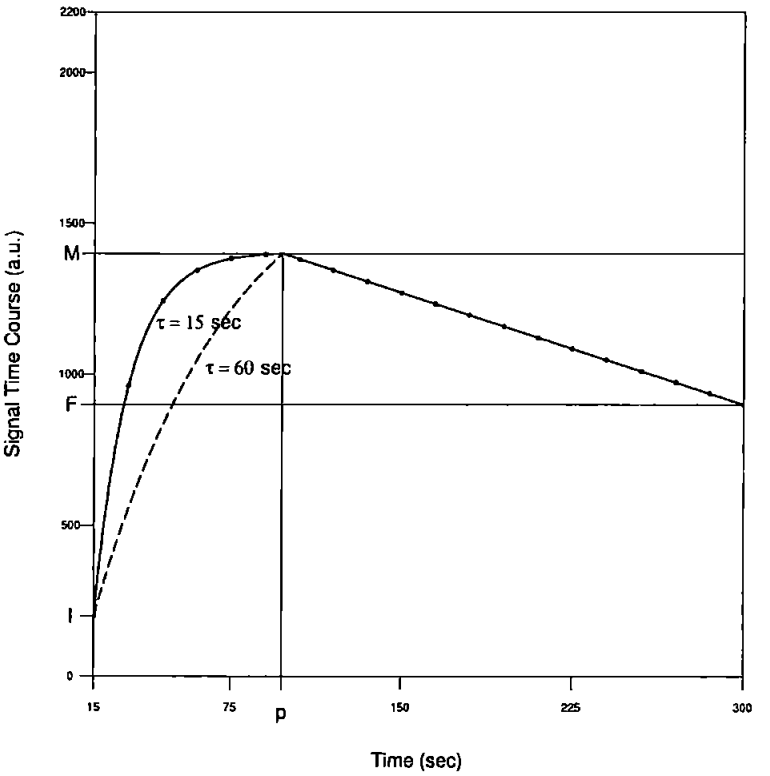


Figure 3.4: *Meaning of the parameters* $\boldsymbol{\theta} = (I, M, F, p, \tau)$ *of the parametric temporal model for the contrast agent uptake. Two parametric time patterns are shown:* $\boldsymbol{\theta} = (200, 1400, 900, 6.5, 15)$ *(solid curve with dots) and* $\boldsymbol{\theta} = (200, 1400, 900, 6.5, 60)$ *(dashed curve). The solid and dashed curves coincide after* p*. The dots indicate the acquired temporal points* $t = 1, \ldots, 20$ *where one time unit is 15 seconds. The units of* τ *are seconds.*

3.3.3 Estimation

In order to minimize the variability between the non-parametric and parametric restorations, in this section we adopt the MMS estimator as used in Section 3.2 to estimate $\boldsymbol{\theta}$. In this way, restoration from the two methods can be compared consistently. To obtain the MMS estimator we adopt the Metropolis-Hastings algorithm with uniform proposal. In fact, in this case we cannot use the modified proposal introduced for the non-parametric case in Section 3.2.3. Furthermore, since we do not have any knowledge

about the distribution of these model parameters a uniform proposal seems the most sensible choice. The parameter ranges for the uniform proposals are defined as follows. The ranges for I and F are the same as those of $\mathbf{y}(1)$ and $\mathbf{y}(T)$ respectively. The range for M corresponds to the minimum and maximum values of the whole measured image data while the range for p is $[2, T-2]$. The range for τ is $\Delta\,[0.2, T/3]$. The MMS estimator is again approximated as the mean of the last half of the Markov chain sequence of images, the length of which is determined by monitoring the spatial mean of the parameters.

3.3.4 Hyper-parameter estimation

In (3.8) there are seven hyper-parameters $(\beta_1, \beta_2, \delta_I, \delta_M, \delta_F, \delta_p, \delta_\tau)$. To estimate δ_I and δ_F we use the procedure described in Section 3.2.4 by considering the empirical distribution of the pixel-wise differences calculated from $\mathbf{y}(1)$ and $\mathbf{y}(T)$. For δ_M, δ_p and δ_τ, we first estimate the images M, p and τ with $\beta_1 = \beta_2 = 0$; these correspond to likelihood based estimates. The values of these three parameters are then set equal to the standard deviation of the empirical distribution of the pixel-wise differences for these images (Glad and Sebastiani, 1995). We cannot use the procedure described in Section 3.2.4 because the relationship between M, p and τ and the observed image sequence does not follow a simple additive model. For the hyper-parameters β_1 and β_2 we follow the approach described in Section 3.2.4.

3.4 Results

3.4.1 Non-parametric approach

In this section we present the results of the hypothesis test based procedure, the hyper-parameter estimation and the consequent restorations. In Fig. 3.5 we compare the

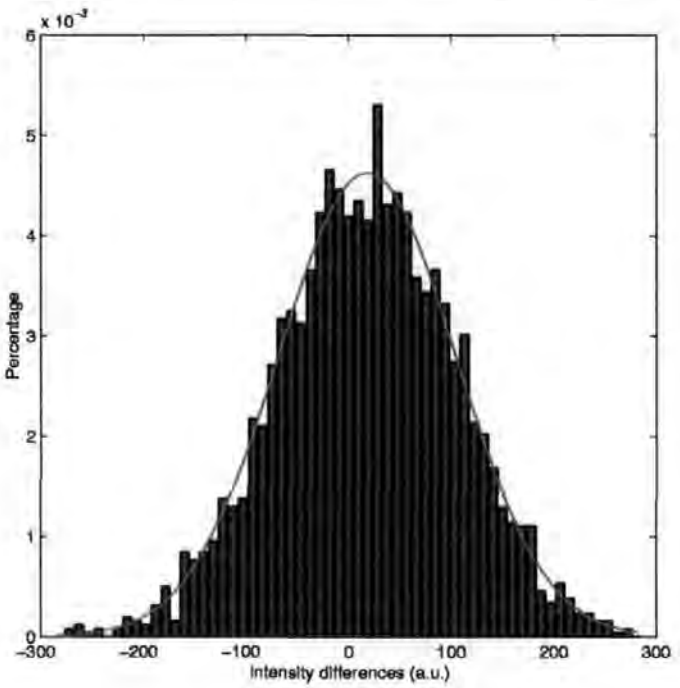


Figure 3.5: *Comparison between the empirical (histogram) and estimated (red line) distributions of pixel values within $\mathcal{R}$. We note a good agreement between these two distributions.*

empirical distribution of pixel values within $\mathcal{R}$ (histogram) and the estimated Gaussian distribution (red line). We note a good agreement between them, so justifying the normality assumption. In Fig. 3.6 we show the results of the hypothesis test corresponding to different values of the significance level α. With $\alpha = 0.50$ we have that 64% of pixels within the ROI are classified tumoral. This percentage reduces to 28% and 21% with $\alpha = 0.10$ and $\alpha = 0.01$ respectively. Therefore, even adopting a very low significance level for the test, we obtain a large percentage of tumoral pixels within the selected ROI.

Furthermore, they all belong to a connected region.

(a) (b) (c)

Figure 3.6: *Results of the hypothesis test corresponding to different significance levels: (a)* $\alpha = 0.10$ *(b),* $\alpha = 0.01$ *(c) and* $\alpha = 0.001$. *In this binary image the bright pixels are tumoral. We note that even with a significance level of 0.01 we still obtain a considerable number of tumoral pixels.*

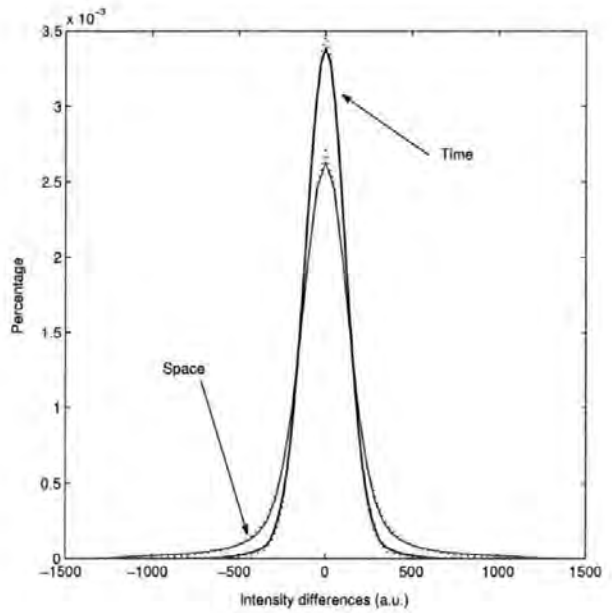
To assess the effectiveness of the procedures for hyper-parameter estimation we show in Fig. 3.7(a) a comparison between $p_l(\Delta y_l)$ for the optimal δ_l and the empirical distribution $\hat{p}_l(\Delta y_l)$ for both space and time. We note a very good agreement between the empirical and theoretical distributions.

To illustrate the results of the adopted criterion for choosing β_s and β_t, we present in Fig. 3.7(b) a comparison between the theoretical χ^2_T distribution and the empirical distribution of $\Sigma_{\mathcal{I}}$ for the optimal β_s and β_t. Again, from this figure we note a good agreement between them. The range used in the optimization procedure and the optimal values are reported in Table 3.1.

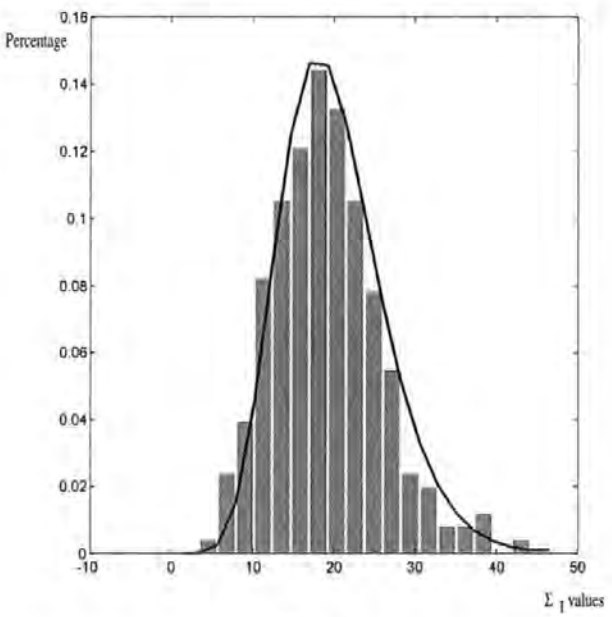
	Range	Optimal value
δ_s	$[10, 100]$	50
δ_t	$[10, 100]$	20
β_s	Not fixed	35
β_t	Not fixed	70

Table 3.1: *Hyper-parameter optimization ranges and the optimal values obtained.*

Now using the optimized hyper-parameter values reported in Table 3.1, we restored



(a)



(b)

Figure 3.7: *(a) Comparison between $p_l(\Delta y_l)$ for the optimal δ_l (continuous line) and the empirical distribution $\tilde{p}_l(\Delta y_l)$ (dotted line) for both space and time. (b) Comparison between the theoretical χ^2_T density (continuous line) and the empirical distribution of $\Sigma_{\mathcal{I}}$ for the optimal β_s and β_t (histogram) for a possible tumoral region $\mathcal{I}$.*

the image sequence. The MMS estimator for the last image of the sequence corresponding to the ROI is shown in Fig. 3.8(b) with the original image being shown in Fig. 3.8(a). In this figure we also show the restoration results obtained for two other patients. In all these restorations we note that the random distortions have been successfully reduced, while the edges of the underlying structure have been preserved, so highlighting information about the tumour morphology.

In order to compare our procedure with a common image restoration method, in Fig. 3.9 we show the restoration obtained by means of a Wavelet based method. In particular, we obtained this result by adopting the Fast Wavelet Transform (FWT) algorithm (Mallat, 1989) for Wavelet decomposition and reconstruction. We tested different Wavelet families and in this figure we report the best result that we obtained. This restoration corresponds to the choice of Symlets Wavelets of order 8 (see Abramovich *et al.*, 2000). In fact, we used three dimensional Symlets, obtained as the tensor product of three one-dimensional Wavelets (first along time, then along row axis and finally along column axis), to restore the images. Following the approach of Donoho and Johnstone (1994, 1995) the significant Wavelet coefficients in the reconstruction were extracted by *thresholding*, i.e. the coefficients are set to zero if their absolute value is below a certain threshold level. We set this threshold level using the same criterion as for the estimation of the hyper-parameters β_s and β_t described in Section 3.2.4. The restoration in Fig. 3.9 was obtained using the estimated optimal value of this threshold level. From the comparison between the Wavelet and Bayesian restorations shown in Fig. 3.8(b) we note that the former is more noisy.

In order to illustrate the restoration effect over time we show in Fig. 3.10 the comparison between the temporal pattern of two pixels within $\mathcal{I}$ before and after the restoration. Here, we note again that the random distortions have been considerably

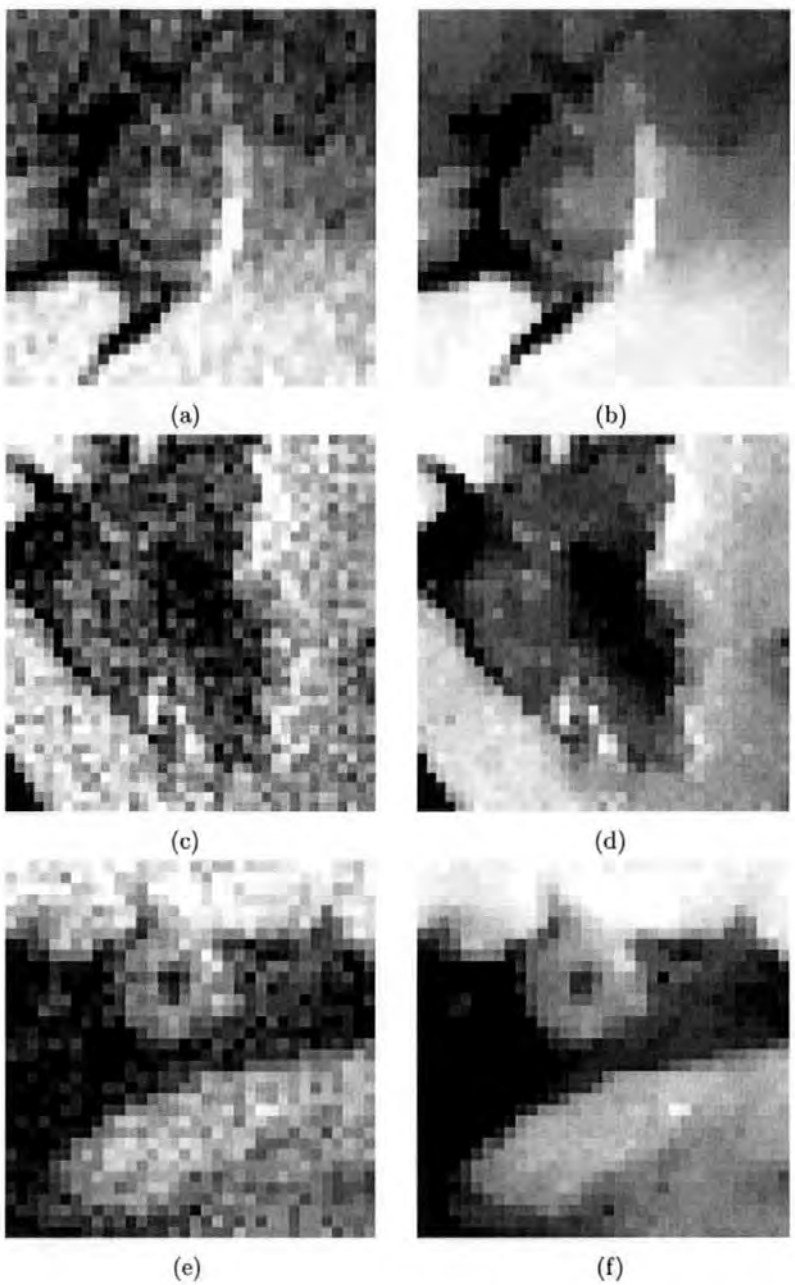


Figure 3.8: *Non-parametric restorations. Comparison between the ROI before (left) and after (right) restoration for three different patients. Here we note how the restoration procedure reduces the random distortions affecting the data and highlights the internal structure of the lesion. In all these cases the last image of the acquired sequence has been shown.*

Figure 3.9: *The restoration obtained using a 3D Wavelet filter.*

reduced allowing us to recognize the typical pattern in Fig. 3.10(a) of benign and in Fig. 3.10(b) of malignant tumoral tissues shown in Fig. 1.1.

3.4.2 Parametric approach

In Fig. 3.11 we show our estimates of the attributes τ and $M - F$ in the tumoral region $\mathcal{I}$. As we can see some structures have become evident. In particular, radiologists involved in this research advise that these ring shaped structures are common in these kinds of tumours. In Fig. 3.12 we show the results of the parametric image restoration procedure for the last image of the sequence in the selected ROI. The restored sequence has been obtained substituting the estimated values of the parameters in the parametric model. This result is very similar to the one obtained from the non-parametric approach shown in Fig. 3.8(b). To assess further the validity of the parametric approach and to compare it with the non-parametric one, we consider the time patterns from the acquired and restored images at two different pixels within $\mathcal{I}$. Fig. 3.13(a) shows an example where the parametric approach provides a better fit than the non-parametric one. This occurs in most of the pixels within the tumoral region, and may lead us to prefer the parametric model. In Fig. 3.13(b) a different situation is presented. Here the data may provide

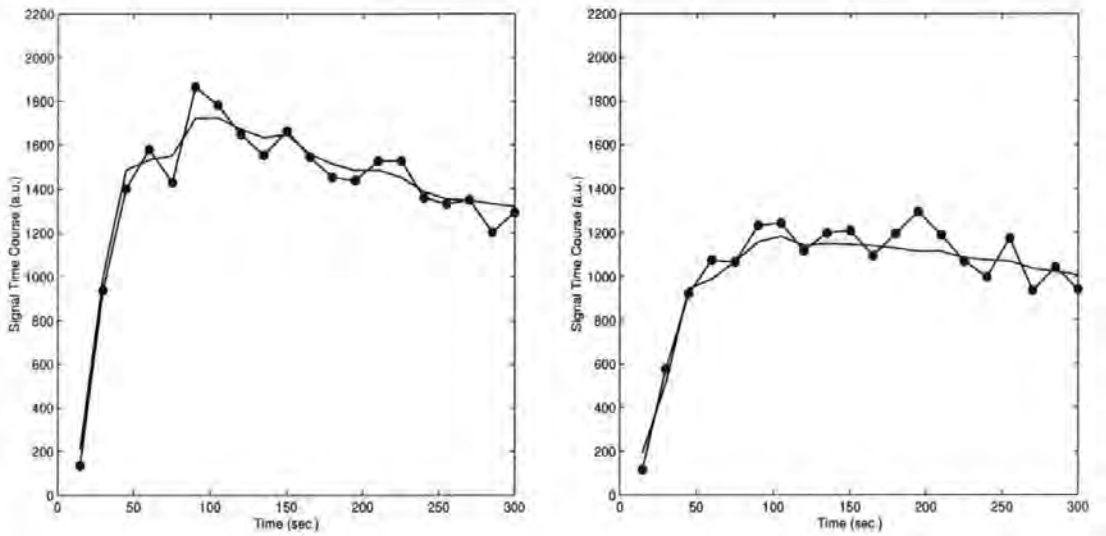
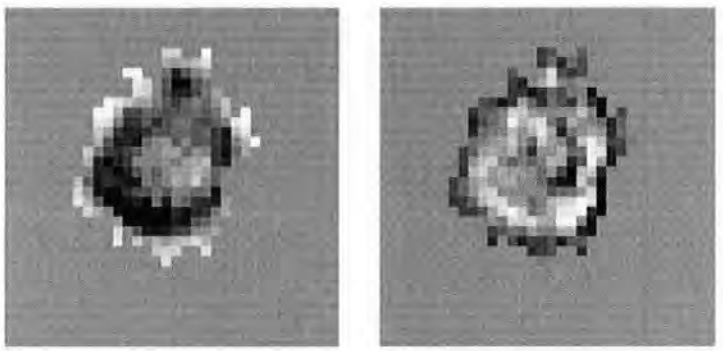


Figure 3.10: *The temporal pattern before (dotted line) and after (continous line) the non-parametric restoration procedure for two pixels within the tumoral region $\mathcal{I}$. In particular the restored patterns in (a) and (b) resemble the expected pattern shown in Fig. 1.1 for malignant and benign tumoral tissue respectively.*

evidence for a time pattern of a different kind than that allowed by the parametric model. In cases like this the parametric model may be too rigid and unable to describe the true temporal evolution properly. Overall, our choice of restoration method should be based on the interpretation of classification results by radiologists. Finally, in contrast to other methods, one of the main advantages of the Bayesian approach is that it can provide us not only point estimates but also credibility intervals for the parameters. This may be useful in further steps of the analysis, such as inter-patient surveys. As an example, in Fig. 3.14 we present the MMS estimator and the associated 95% credibility intervals for a central row of the parameter τ shown in Fig. 3.11.



τ $M - F$

Figure 3.11: *Images of the estimated attributes τ and $M - F$ from the parametric approach in the tumoral region $\mathcal{I}$.*

Figure 3.12: *Parametric restoration. As in Fig. 3.8 we show the last image of the sequence. This restoration is very similar to the non-parametric one shown in Fig. 3.8.*

3.5 Conclusions

A disadvantage of MCMC methods for analysing image sequences is the high computational time that they require due to the typically large number of iterations needed to reach convergence. In our case, since the dMRI sequence consists of 20 images of 256×256 pixels, it is impossible to perform the analysis on the whole data set in the time that is typically available for these kinds of studies. In fact, to make our approach appealing we need the results in a few minutes. For this reason, the first step of identifying

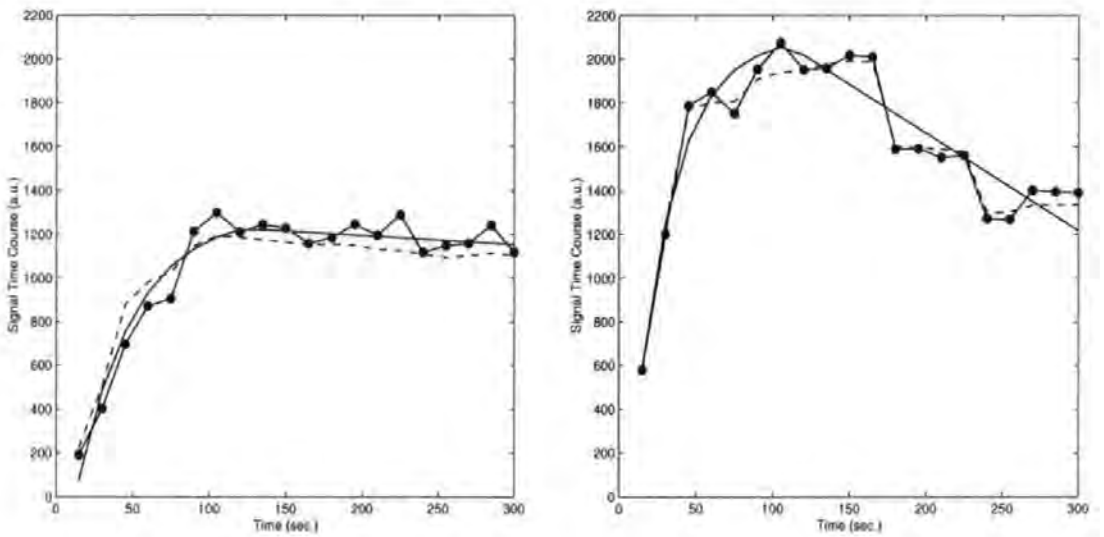


Figure 3.13: *Comparisons between the acquired image intensity time profiles (line with dots) and the restored profile from the non-parametric (dashed line) and the parametric (continuous line) approach. The profiles in (a) and (b) correspond to different pixels. One time unit is 15 seconds.*

the tumoral region $\mathcal{I}$ within the breast is very important in reducing the computational burden. The results show that our hypothesis test based method gives us a small reliable region even if we choose a very low significance level such as 1%. Restricting attention to a small tumoral region reduces the number of pixels analysed from approximately a million to few thousands.

To speed up further our procedure we developed a single component M-H algorithm with specially chosen proposal function based on the data distribution. In particular, the proposal is taken to be normal with expected value equal to the acquired data and variance set to the value estimated from the image intensity in the background, as explained in Section 3.2.2. Comparisons of the performance of the M-H algorithm with different proposals, namely the optimal Gaussian random walk, a uniform independence proposal and our proposal, show how our algorithm reaches convergence earlier than the others.

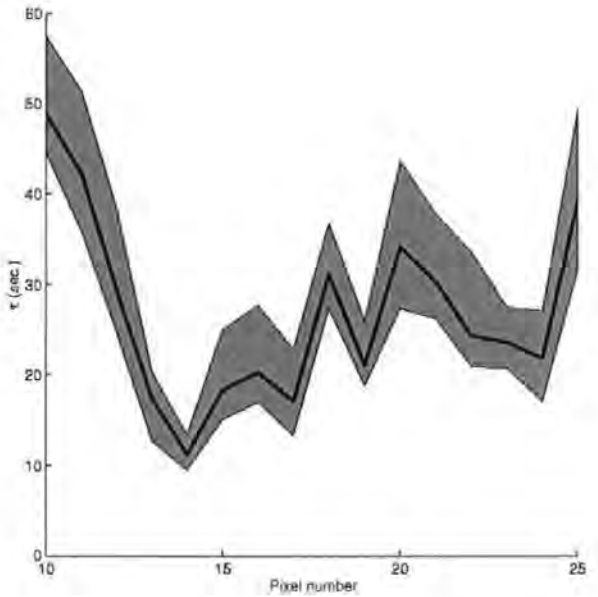


Figure 3.14: *Point estimates and 95% credibility intervals for a central row of the parameter τ shown in Fig. 3.11*

Another issue addressed in this chapter is the estimation of the hyper-parameters of the prior models. We propose criteria to estimate these parameters making our procedure automatic. In this way our methods can be used by non-experts to analyse large samples of patients, thus providing a validation for the developed techniques.

Both non-parametric and parametric image restorations show that the estimated hyper-parameter values are reliable. In fact, in the restored image sequences random distortions have been successfully reduced while the underlying structures have been preserved. The restorations obtained using the two approaches seem similar. However, the temporal patterns obtained from the two methods show some differences. In some cases the non-parametric approach provides a biased estimate of the temporal evolution, while in other cases the parametric approach seems too rigid and unable to describe the temporal evolution properly. Our analysis will be mainly focused within the identified

tumoral region $\mathcal{I}$ where the parametric approach seems generally to perform better.

Finally we remark that one of the advantages of adopting a Bayesian approach compared to other methods is that it can provide not only point estimates but also credibility intervals. In this chapter, we show the 95% point-wise credibility interval for a central row of the ROI for the attribute τ of the parametric model. This information could be useful in further steps of our analysis such as inter-patients surveys.

Chapter 4

Bayesian Image Classification

In this chapter we describe our Bayesian image classification step. This step is based on the restored dMRI sequence and the method that we developed is motivated by the work of Sebastiani and Sørbye (2002). The classification is based on a few attributes describing the relevant features of the image intensity time pattern for each pixel. Initially, we tested different choices for these attributes, for both the non-parametric and parametric approach; the results that we present are obtained by adopting those that performed best. In particular, from the non-parametric approach, we adopted the areas A and B shown in Fig. 4.1. These areas quantify gadolinium wash-in and wash-out and are easily calculated from the smoothed temporal patterns. We chose to adopt integral-based attributes because they are more robust than attributes defined as the difference between two points of the temporal pattern of the signal. Furthermore, they include all the acquired temporal information for every pixel. From the parametric approach we use the parameters τ and $M - F$, as these again quantify accurately gadolinium wash-in and wash-out. For comparable values of $M - I$, A increases with τ. However, although we could calculate A from the estimated temporal model, doing so would result in less spatial regularity than using τ for which we adopt a spatial prior model.

Let $\mathbf{d_i} = (\mathrm{d}_{1,i}, \ldots, \mathrm{d}_{m,i})$ be a vector containing the m attributes at pixel i. In our

case we take $m = 2$. The classification procedure can be represented by a mapping of every pixel i to a class $k_i \in \{1, 2, \ldots, K\}$ where K is the number of classes considered. The value of K can be chosen by the radiologist. If the classification is performed within a selected ROI, we set $K = 3$, corresponding to normal, benign and malignant tumoral tissues. Alternatively, if the procedure is applied within the tumoral region $\mathcal{I}$ described in Section 3.2.4, we set $K = 2$ corresponding to benign and malignant tumoral tissues. In this study we restrict our classification procedure within $\mathcal{I}$. Thus our procedure tries to identify different physiopathological conditions of the tumoral tissues. Following the

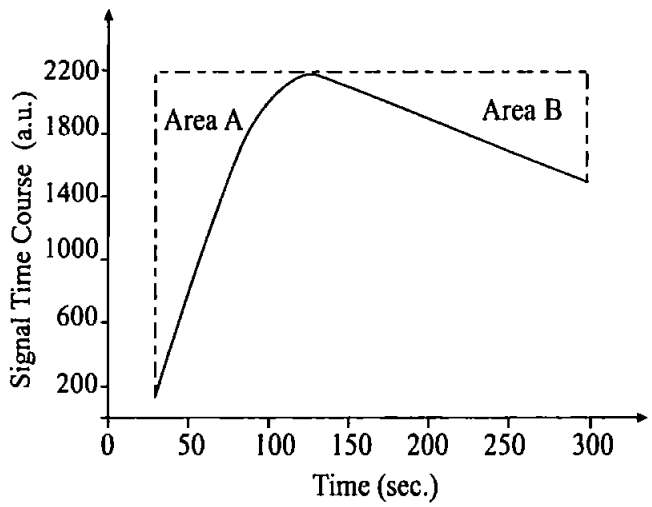


Figure 4.1: *Classification attributes for a given pixel are taken to be the areas A and B.*

scheme of Chapter 3, in Section 4.1 and 4.2 we introduce the attribute and the prior model, while in Section 4.3 we describe how the estimation is performed. Finally, in Section 4.4 the classification results obtained from both the non-parametric and parametric attributes are shown. These results are compared with those currently available to radiologists for locating the lesion and with the classification obtained from a standard hierarchical cluster analysis.

4.1 Attribute model

The j^{th} attribute for the k^{th} class is assumed to follow a Gaussian distribution with expected value $c_{j,k}$ and variance $V_{j,k}$. The attributes are assumed to be conditionally independent with the result that the distribution of $\mathbf{d} = (\mathbf{d}_1, \ldots, \mathbf{d}_m)$ given the classification vector $\mathbf{k} = (k_1, \ldots, k_n)$ is:

$$p(\mathbf{d}|\mathbf{k}, \mathbf{c}, \mathbf{V}) = \prod_{i=1}^{n} \prod_{j=1}^{m} (2\pi V_{j,k_i})^{-1/2} \exp\left\{-(d_{j,i} - c_{j,k_i})^2/(2V_{j,k_i})\right\},$$

where $\mathbf{c} = (\mathbf{c}_1, \ldots, \mathbf{c}_K)$ and $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_K)$, in which $\mathbf{c}_k = (c_{1,k}, \ldots, c_{m,k})$ and $\mathbf{V}_k = (V_{1,k}, \ldots, V_{m,k})$ are the attribute means and variances for tissue class k. These vectors are assumed unknown and will be estimated at the same time as $\mathbf{k}$.

4.2 Prior model

Our prior assumption about the classified image is that neighbouring pixels are more likely to belong to the same class than to different classes. Hence we adopt the Potts model (Potts, 1952):

$$\mathrm{P}(\mathbf{k}|\beta) = \frac{1}{Z_\beta} \exp(-\beta U_{\mathbf{k}}) = \frac{1}{Z_\beta} \exp\left(\beta \sum_{<ij>} \delta_{k_i,k_j}\right),$$

where the hyper-parameter β has to be estimated and Z_β is the unknown normalizing constant.

4.3 Estimation

Inference about $\mathbf{k}, \mathbf{c}, \mathbf{V}$ and β is based on their joint posterior distribution. By using Bayes theorem and the reasonable simplifying assumptions that $\mathrm{P}(\mathbf{d}|\mathbf{k}, \mathbf{c}, \mathbf{V}, \beta) = \mathrm{P}(\mathbf{d}|\mathbf{k}, \mathbf{c}, \mathbf{V})$ and that $\mathrm{P}(\mathbf{k}|\beta, \mathbf{c}, \mathbf{V}) = \mathrm{P}(\mathbf{k}|\beta)$, we obtain $\mathrm{P}(\mathbf{k}, \mathbf{c}, \mathbf{V}, \beta|\mathbf{d}) \propto \mathrm{P}(\mathbf{d}|\mathbf{k}, \mathbf{c}, \mathbf{V})\,\mathrm{P}(\mathbf{k}|\beta)\,\mathrm{P}(\mathbf{c}, \mathbf{V}, \beta)$, where $\mathrm{P}(\mathbf{c}, \mathbf{V}, \beta)$ is the prior

distribution on $\mathbf{c}$, $\mathbf{V}$ and β. Since we do not have any knowledge about the distribution of these parameters, here we assume a uniform prior in a suitable range. As we now adopt a fully Bayesian approach (Besag, 1989), we need to estimate Z_β up to a proportionality constant for a range of values of β. Since $Z_\beta = \sum_{\mathbf{k}} \exp(-\beta U_{\mathbf{k}})$, as we mentioned in Section 1.4.5, exact calculation is not feasible due to the high number K^n of configurations involved. However, we proceed by noting that:

$$\frac{\partial \log Z_\beta}{\partial \beta} = \frac{1}{Z_\beta}\frac{\partial Z_\beta}{\partial \beta} = -\sum_{\mathbf{k}} \frac{1}{Z_\beta} U_{\mathbf{k}} \exp(-\beta U_{\mathbf{k}}) = -E_{\mathrm{P}(\mathbf{k}|\beta)}[U_{\mathbf{k}}], \tag{4.1}$$

where $E_{\mathrm{P}(\mathbf{k}|\beta)}[U_{\mathbf{k}}]$ is the expected value of the energy function $U_{\mathbf{k}}$ under $\mathrm{P}(\mathbf{k}|\beta)$. Then, by integrating (4.1) with respect to β we obtain:

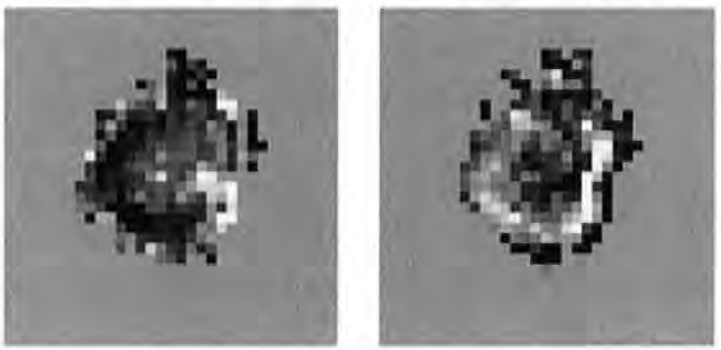
$$\log(Z_\beta) - \log(Z_{\beta_0}) = -\int_{\beta_0}^{\beta} E_{\mathrm{P}(\mathbf{k}|\beta')}[U_{\mathbf{k}}]\, d\beta',$$

for some fixed value β_0. We approximate $E_{\mathrm{P}(\mathbf{k}|\beta')}[U_{\mathbf{k}}]$ using the M-H algorithm for a finite number of values of β' and calculate the integral numerically using Simpson's rule. Related approaches for approximating the normalizing constant can be found in Gelman and Meng (1998) and Green and Richardson (2002).

Our chosen estimator is the MAP estimator, the classification image corresponding to the maximum of the posterior distribution. The MAP estimator is discussed in detail in Section 1.4.4 and 1.4.5. To decrease the computational burden we obtain an approximation of the MAP estimator by using the Iterative Conditional Mode (ICM) algorithm presented in Besag (1986) and outlined in Section 1.5.2. As the ICM algorithm is a local optimization procedure, it depends on its starting point. We chose the K-means clustering algorithm output as the starting point. This algorithm is discussed by Glasbey and Horgan (1995) and outlined in Section 1.5.1. The initial point for the K-means clustering algorithm is a classification image based on the modes of the histogram of one of the class attributes.

4.4 Results

In Fig. 4.2 we show the non-parametric classification attributes A and B within $\mathcal{I}$. Structures resembling a 'C' and a 'U' seem to appear in this figure. Although these structures are not clearly defined, the information in the A and B images will lead the classification procedure towards a meaningful structure as we will now see. In order to



(a) (b)

Figure 4.2: *Non-parametric classification attributes A in (a) and B in (b). Structures resembling a 'C' and a 'U' seem to appear in (a) and (b) respectively.*

assess the performance of the methodology that we developed, we begin by considering the results obtained using two existing techniques. First, in Fig. 4.3(a) we show the result that is obtained using the method currently employed by radiologists. This method is based on the mean difference image $\mathbf{m}$ introduced in equation (3.1) and only identifies the location of the lesion. The lesion location, obtained by extracting the contour of the bright region within $\mathbf{m}$ shown in Fig. 3.1(b), is superimposed in red on the last image of the sequences.

An analysis of the temporal patterns of pixels inside this red area confirmed that it corresponds to a tumoral region. Hence, this method can provide a clear indication of the tumoral lesion location although it provides no information about tumour morphology.

Next, we consider the image classification obtained by applying a standard statistical

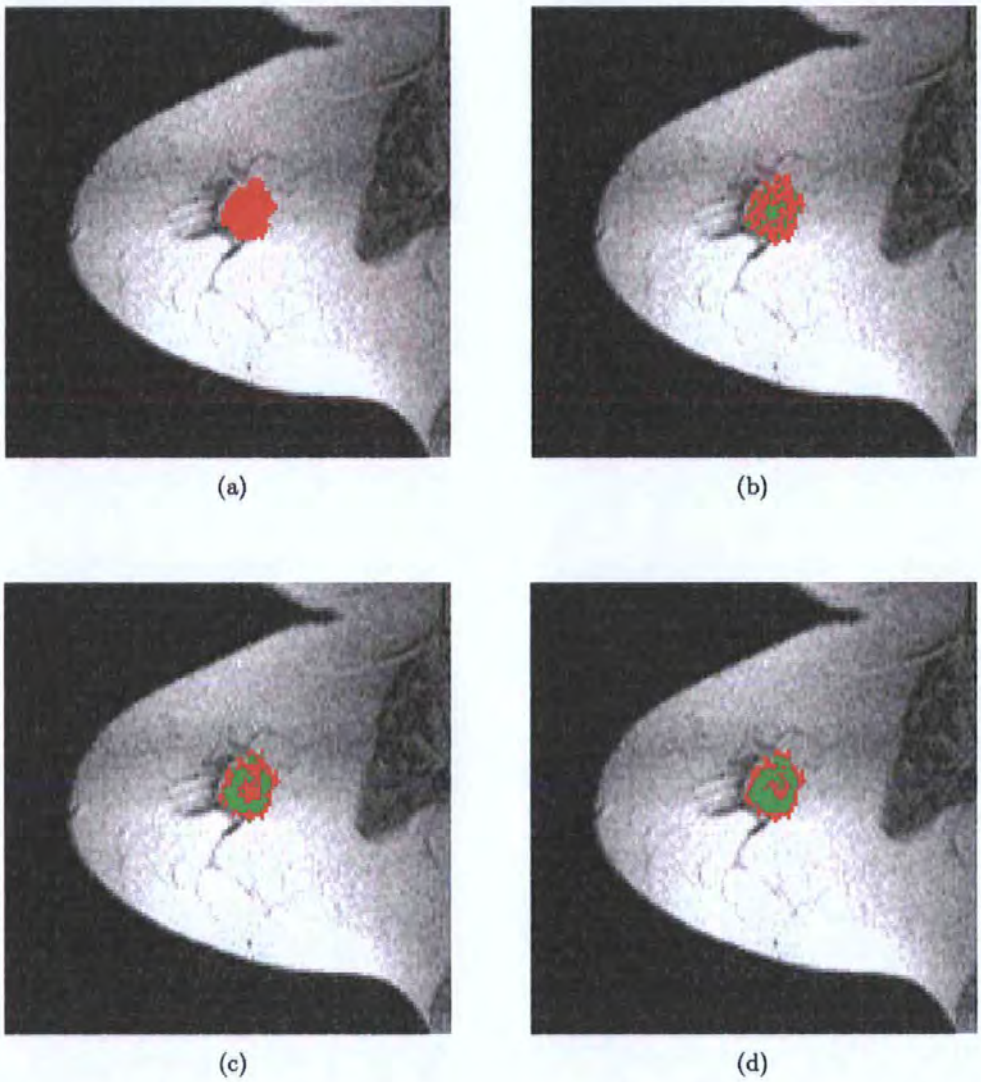


Figure 4.3: *(a) The lesion location obtained from the method currently employed by radiologists based on the mean difference image* $\mathbf{m}$. *(b) Hierarchical clustering classification. The two clusters are identified by red and green. This classification is very speckled and no internal structure is evident. (c) Non-parametric classification. A structure resembling a 'C' appears clearly indicating a necrotic region (green) surrounded by a more active region (red). (d) Parametric classification. A ring shaped structure clearly appears. Similar considerations to (c) about the morphology of the tumour apply.*

method to the raw dataset. This is because we did not want this procedure to be influenced by the Bayesian restoration step. The adopted technique is a hierarchical clustering method (Krzanowski, 1990). In order for the results not to be biased by the choice of classification attributes, we used as data the whole temporal pattern, i.e. all twenty temporal values at every pixel. Among the different choices that we considered, we obtained the best result using the Mahalanobis distance matrix and the 'furthest' linkage criterion. Since the classification was performed within the tumoral region $\mathcal{I}$, the number of clusters was fixed to two. In Fig. 4.3(b) we show the clusters that we produced. As we can see, the classification is very speckled and no internal structure of the tumour is evident.

To illustrate the improvements that can result from our Bayesian methods, the classification image based on the non-parametric attributes A and B is shown in Fig. 4.3(c). In particular, this result has been obtained by classifying the pixels within region $\mathcal{I}$ into two classes: benign (red) and malignant (green) tumoral tissues. We note the presence of a 'C' structure that is frequently present in these kinds of tumours (Heywang-Kobrunner and Beck, 1995). Radiologists have advised us that this structure is consistent with the presence of a necrotic region at the centre of the tumour caused by the tumoral growth mechanism.

In Fig. 4.3(d) we show the classification based on the attributes τ and $M-F$ obtained from the parametric approach. Very similar results were achieved by using the parameters p and $M-F$. We note the presence of a ring structure, again compatible with the tumoral nourishment mechanism.

In addition, in order to check the temporal pattern of the pixels belonging to the two classes, we averaged the temporal evolution of the signal over the classified regions for the classification image obtained from the non-parametric attributes. In Fig. 4.4 we

show these mean temporal patterns for the two classes. We note that they resemble the patterns discussed in Section 1.2. Almost identical curves were obtained from the parametric classification presented in Fig. 4.3(d). The above results further validate the success of the whole procedure for investigating the internal structures of the lesion.

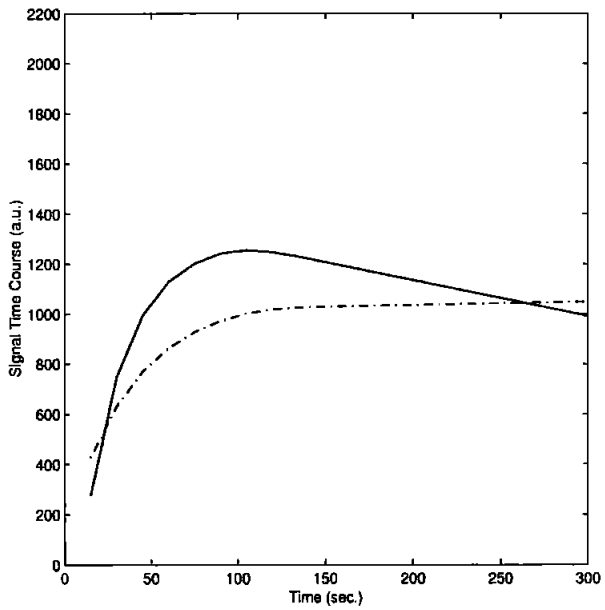


Figure 4.4: *The mean temporal patterns over the two classes. We note a good agreement with the characteristic patterns for malignant and benign tumoral tissues shown in Fig. 1.1.*

4.5 Conclusions

The classification of tissues inside a tumoral lesion is very important both to understand the way in which the lesion is growing and to plan different medical interventions. It must be stressed that dMRI can provide information about the tumour morphology before any medical intervention and hence before the tumour is removed. The method that radiologists currently use to analyse these sequences can provide an indication of the tumour location, but no information about the internal structure of the tumour can be gathered. In order to show the improvement obtained by our methodology we first

performed image classification using a hierarchical clustering method. This was applied to the raw data using as pixel attributes the whole temporal pattern. In this way, the results were not biased by our image restoration and our choice of classification attributes. The classification obtained clearly shows that such a procedure is not able to identify any structure within the lesion. This is due to the noise that affects the data. Our Bayesian classification, in which the image sequence has been previously restored, prior knowledge has been modelled and attributes have been chosen carefully, seems very promising in both the non-parametric and parametric case. In fact, our results show a clear improvement over those currently available: not only do the tumoral regions identified by our method and the technique currently used by radiologist coincide, but also an underlying internal lesion structure becomes evident from our classification image. This indicates that our procedure seems to be able to discriminate between different pathological conditions of the tumoral tissues. This result is very important from a clinical point of view. At this stage we are not able to say whether the non-parametric or parametric approach produces the more reliable estimate of the lesion structure. Our experience is that the results obtained from the two methods are similar and both structures have been judged to be reliable by radiologists. A further indication that the classification images are accurate, at this stage, can be obtained by considering the mean temporal patterns inside the two classes identified within the lesion. As an example, we show these mean patterns for the non-parametric classification. We note that they resemble very closely the expected theoretical patterns shown in Fig 1.1.

Chapter 5

Deformable Template Models

In this chapter we present an image classification method that, unlike the Bayesian procedure introduced in Chapter 4, takes into account the spatial structure of the lesion. In fact, as we mentioned in Section 1.2, tumour morphology can be a very useful classification criterion. The approach that we shall develop will exploit knowledge about the shape of lesions that radiologists often have before acquiring the data. Examples of such shape information are given in Fig. 1.2. In our Bayesian classification methodology the spatial structure of the lesion is taken into account by prior distributions modelling spatial homogeneity and temporal continuity between neighbouring pixels of the image sequence. These models try to minimize random distortions affecting the data while preserving the edges that characterize the tumour contour. No high level priors modelling the actual shape of the lesion were adopted in the Bayesian methodology. However, in the approach described in this chapter knowledge about the shape of the tumour is taken into account by means of deformable template models.

Our new procedure is based on the classification attributes introduced in Chapter 4, restricted to a previously identified region $\mathcal{I}$, such as the one discussed in Section 3.1 and shown in Fig. 3.6(c). In this way, our task will be to identify the border separating the malignant and benign tumoral classes within the lesion. We assume that the border

separating these classes is a simply connected line so that it can be represented as a shape in a binary image. The pixels inside (outside) this shape will be classified as malignant (benign) tumoral tissue. Throughout this chapter for simplicity we refer to the malignant tumoral tissue as a lesion, so that by the lesion contour we mean the border separating the two classes. In this way, reconstructing the lesion contour is equivalent to obtaining a classification image.

The basic idea is that we start from a template in the form of an image that represents our knowledge about the lesion shape. We then deform this template using parametric transformations to produce an image that matches the lesion contour. This image is obtained by minimizing an objective function that depends on the parameters of the deformation. Hence, our matching algorithm consists of the following three steps:

1. introducing a prototype template;
2. deforming the prototype by using a set of parametric transformations;
3. minimizing an objective function over the set of parametric transformations in order to match the deformed template to the lesion contour.

In Section 5.1 we describe our choice of prototype template. In Section 5.2 we derive the parametric transformation model and discuss the meaning of its parameters. In Section. 5.3, we investigate the richness of our deformation models by means of a simulation study. In Section 5.4 we introduce the objective function, while in Section 5.5 we present full details of our matching algorithm. In Section 5.6 we present our results. These were obtained from a simulation study in which we investigated the potential of our algorithm by applying it to several datasets with different Signal to Noise Ratios (SNR). We also applied our methodology to our real dataset with success.

5.1 The prototype template

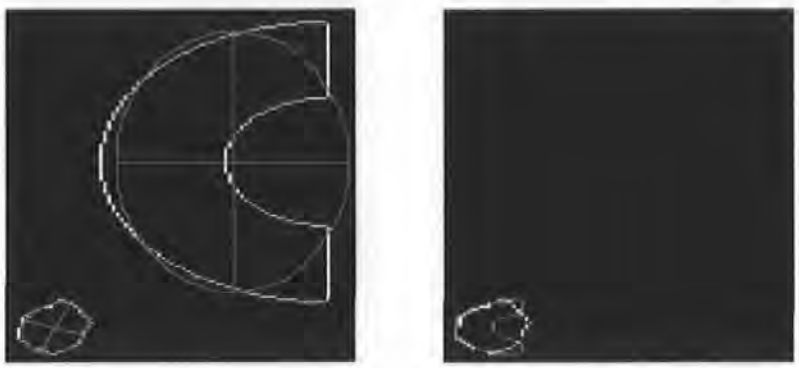
The prototype template is a binary image that represents our prior knowledge about the shape of the lesion. This can be obtained from a training set as in the 'Eigenshapes models' discussed by Baumberg and Hogg (1995), or it can be drawn by expert radiologists. As mentioned in Section 1.5, a prototype template that can be deformed by parametric mappings has the advantage over a parameterized template that many different shapes can be very easily included in the analysis. Furthermore, depending on how accurate our prior knowledge about the tumour shape is, detailed or coarse information about the contour separating the two classes can be modelled. For example, if for a particular lesion we do not have any specific information about the shape, we can take the contour of the tumoral region $\mathcal{I}$ as the prototype template and allow large deformations in the matching procedure. On the other hand, if we know that the lesion under investigation resembles a particular shape, we can specify the shape and allow only very small deformations during the analysis. An important feature of this approach is that the distance between the estimated final shape and the prototype template can be quantified by the parameters of the deformation and this information can be very useful from a clinical point of view. In fact, if the prototype image represents a particular pathology, this distance can represent how accurately the lesion can be classified as belonging to that pathology.

Since the prototype represents the initial shape of the matching algorithm, it plays a very important role. Our 'a priori' knowledge about the tumoral shape consists of the lesion contour, size and location within the breast. At this stage, we assume that the lesion border is obtained from an expert radiologist who draws the contour that he expects to find for the kind of tumours under study. This contour is shown as the white 'C' shape in Fig. 5.1(a). The size and location of the lesion are roughly estimated by means of the

tumoral region $\mathcal{I}$ shown in Fig. 3.6(c). These estimates are used to scale and shift the prototype. We now describe how this step is performed. Initially the contour of region $\mathcal{I}$ is extracted and its centroid is calculated. Since we assume that the centroids of the tumoral region and lesion are the same, we translate the prototype in such a way that these two centroids coincide. Once the prototype is centred, we perform the scaling. In order to estimate the scaling factors we adopted a method based on the best fitting ellipses for the contour of $\mathcal{I}$ and the shifted prototype. In particular, the ellipses are obtained by matching the moments of these shapes. This step is described in detail in Appendix A. In Fig. 5.1(a) we show the best fitting ellipses superimposed on the $\mathcal{I}$ contour and the prototype template before translation. In order to scale the prototype, the lengths of the major and minor axes of the ellipses are calculated. Now, the prototype is scaled along the horizontal and vertical axis by factors given by the ratio of the lengths of the two major and minor axes respectively. Since the tumoral region contour does not contain any information about the lesion orientation, we do not apply any rotation to the prototype at this stage. Rotation of the prototype, as will become clear in Section 5.2, can be approximately obtained by combining different transformations given by the adopted deformation model. The shifted and scaled prototype template (red) is shown in Fig. 5.1(b) with the $\mathcal{I}$ contour. Throughout our study we will use the 'C' shape as prototype contour and translate and scale it in the way we have described. This translated and scaled prototype will be the initial point for our matching algorithm and will be denoted by $\mathbf{T}_0$.

5.2 The deformation model

Once the translated and scaled prototype $\mathbf{T}_0$ has been obtained it will be deformed by using a set of non-affine transformations. We now describe these transformations. An intuitive idea of the effect of these non-affine deformations can be provided by employing



(a) (b)

Figure 5.1: *(a) The white 'C' shape is the contour supplied by the radiologist. The contour of region $\mathcal{I}$ is also shown in white. The best fitting ellipses (green) are superimposed on those contours. (b) The contour supplied by the radiologist is shifted and scaled to obtain the prototype template* $\mathbf{T}_0$ *which is superimposed (red) on the* $\mathcal{I}$ *contour.*

a rubber sheet representation. We can assume that the template is drawn on a planar rubber sheet with fixed boundaries lying in the unit square $\mathbf{S} = [0,1]^2$ and that this sheet can be deformed by locally stretching, squeezing and twisting the interior. As the rubber sheet deforms, the template on it changes shape. Such non-affine transformations can be described by a continuous mapping $\mathcal{K}$ of $\mathbf{S}$ into itself:

$$(x,y) \overset{\mathcal{K}}{\mapsto} (x,y) + \mathbf{U}(x,y), \tag{5.1}$$

where $\mathbf{U}(x,y) = (U^1(x,y), U^2(x,y))$ is a parametric transformation. To specify $\mathbf{U}(x,y)$ we can represent the rubber sheet deformation by a partial differential equation that in physics represents the movements of a membrane subject to the application of a force $\mathbf{f}$ at a given point (x,y):

$$\mathcal{L}^2(\mathbf{U}) = \mathbf{f}, \tag{5.2}$$

where

$$\mathcal{L}^2(\mathbf{U}) = \left(\frac{\partial^2 U^1}{\partial x^2} + \frac{\partial^2 U^1}{\partial y^2}, \frac{\partial^2 U^2}{\partial x^2} + \frac{\partial^2 U^2}{\partial y^2} \right).$$

If we adopt the following boundary conditions of mixed type for (5.2):

$$\begin{array}{ll} U^1(0,y) = U^1(1,y) = 0; & U^2(x,0) = U^2(x,1) = 0 \\ U^1_y(x,0) = U^1_y(x,1) = 0; & U^2_x(0,y) = U^2_x(1,y) = 0, \end{array} \tag{5.3}$$

the borders are mapped into themselves in such a way that abrupt transformations are avoided near them. To find the solution of (5.2) subject to the boundary conditions (5.3) is straightforward. In fact, if we know the Green's function $\mathbf{G}(x,y;\xi,\eta) \equiv (G^1(x,y;\xi,\eta), G^2(x,y;\xi,\eta))$ that satisfies both partial differential equations:

$$\frac{\partial^2 G^i}{\partial x^2} + \frac{\partial^2 G^i}{\partial y^2} = \delta(x-\xi)\,\delta(y-\eta), \qquad i = 1,2 \tag{5.4}$$

where δ is the Dirac delta distribution and $(\xi,\eta) \in \mathbf{S}$, and the boundary conditions (5.3), then $\mathbf{U}(x,y)$ can be calculated by:

$$\mathbf{U}(x,y) = \int_0^1 \int_0^1 \mathbf{G}(x,y;\xi,\eta) \otimes \mathbf{f}(\xi,\eta) d\xi d\eta, \tag{5.5}$$

where multiplication $\otimes$ and integration are applied component-wise. Now, it can be shown that $\mathbf{G}$ can be expressed in terms of the eigenvalues λ_{nm} and eigenvectors $\psi_{nm} = (\psi^1_{nm}, \psi^2_{nm})$ of $\mathcal{L}^2$ through the bilinear formula as (Butkov, 1968):

$$\mathbf{G}(x,y;\xi,\eta) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{\psi_{nm}(\xi,\eta) \otimes \psi_{nm}(x,y)}{\lambda_{nm}}, \tag{5.6}$$

where summation is applied component-wise. With the boundary conditions (5.3), ψ_{nm} represent an orthonormal basis, the components of which are given by:

$$\begin{aligned} \psi^1_{nm}(x,y) &= 2\sin(n\pi x)\cos(m\pi y) \\ \psi^2_{nm}(x,y) &= 2\cos(m\pi x)\sin(n\pi y) \end{aligned} \tag{5.7}$$

both of which have associated eigenvalues $\lambda_{nm} = -2\pi^2(n^2+m^2)$. Substituting (5.7) into (5.6), we obtain:

$$G^1(x,y;\xi,\eta) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{1}{\pi^2(n^2+m^2)} [-\sin(n\pi\xi)\cos(m\pi\eta)]\, 2\sin(n\pi x)\cos(m\pi y)$$

$$G^2(x,y;\xi,\eta) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{1}{\pi^2(n^2+m^2)} \left[-\cos(m\pi\xi)\sin(n\pi\eta)\right] 2\cos(m\pi x)\sin(n\pi y)\,. \tag{5.8}$$

We now assume that $\mathbf{f}$ is an impulsive force at a point $(\gamma,\zeta) \in \mathbf{S}$, the components of which can be represented by:

$$f^i(x,y) = \frac{1}{\alpha}\delta(x-\gamma)\delta(y-\zeta), \quad i = 1,2, \tag{5.9}$$

where $\alpha = \dfrac{F}{A}$, in which F represents the tension in the membrane and A the external load per unit area. Substituting (5.9) into (5.5), we have:

$$\mathbf{U}(x,y) = \frac{1}{\alpha}\,\mathbf{G}(x,y;\gamma,\zeta);$$

so from (5.6) we obtain

$$\mathbf{U}_{\alpha;\boldsymbol{\varepsilon}}(x,y) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{1}{\alpha\pi^2(n^2+m^2)}\,\boldsymbol{\varepsilon}_{nm} \otimes \psi_{nm}(x,y), \tag{5.10}$$

where $\boldsymbol{\varepsilon}$ is a set containing all the parameters $\boldsymbol{\varepsilon}_{nm}$ and $\boldsymbol{\varepsilon}_{nm} = (\varepsilon^1_{nm}, \varepsilon^2_{nm})$ is a vector of deformation parameters with components given by:

$$\begin{aligned} \varepsilon^1_{nm}(\gamma,\zeta) &= -\sin(n\pi\gamma)\cos(m\pi\zeta) \\ \varepsilon^2_{nm}(\gamma,\zeta) &= -\cos(m\pi\gamma)\sin(n\pi\zeta)\,. \end{aligned}$$

By varying the application point (γ,ζ) of the force $\mathbf{f}$ these deformation parameters will range in $[-1,1]^2$. Clearly the form of the deformation parameters derives from the particular choice of $\mathbf{f}$. Different choices of the force lead to different forms and hence ranges for $\boldsymbol{\varepsilon}_{nm}$ (see Amit *et al.*, 1991 and Jain *et al.*, 1996). In order to widen the class of deformations that we consider, we shall assume that ε^1_{nm} and ε^2_{nm} can take any value in $[-1,1]^2$ independently of each other.

The parameter α plays an important role in our deformation models. This parameter can be thought of as a scaling factor as it controls the amplitude of the deformation. The physical interpretation is that for a given value of the external load A, α is proportional to the membrane tension F. If F increases, the membrane becomes more rigid so that the same load produces smaller deformations in every point of the membrane. Once the values of ε_{nm} and α are fixed, the transformation is uniquely determined.

In our implementation, because only a finite number of terms in the infinite series (5.10) can be used, the deformation field is approximated by:

$$\mathbf{U}_{\Theta}(x,y) = \sum_{n=1}^{N}\sum_{m=1}^{M} \frac{1}{\alpha\pi^2(n^2+m^2)}\,\varepsilon_{nm} \otimes \psi_{nm}(x,y), \tag{5.11}$$

where $\Theta = (\alpha; \varepsilon; N, M)$, and ε.

By setting the number of components N and M to large values it is possible to obtain very complex deformations. However, in this study we keep N and M small because we do not want to generate shapes that are very different from the prototype template. In addition, if we increase the number of components, the number of deformation parameters increases. Thus, the computational burden associated with the optimization of the objective function, that will be introduced in Section 5.4, becomes too onerous. For these reasons, we set $N = M = 1$. For notational simplicity we henceforth write $\Theta = (\alpha; \varepsilon)$ in which $\varepsilon = (\varepsilon^1, \varepsilon^2)$.

Returning to (5.1), we can now define the mapping operator as:

$$\mathcal{K}_{\Theta}(x,y) = (x,y) + \mathbf{U}_{\Theta}(x,y).$$

For a binary image $\mathbf{X}_0$ with pixel centres indexed by (x,y), we shall let $\mathbf{X}_1 = \mathcal{K}_{\Theta}\mathbf{X}_0$ be the image that results from transforming each pixel of $\mathbf{X}_0$ by $\mathcal{K}_{\Theta}$.

We now show the deformation of $\mathbf{S}$ for different choices of Θ. To do this, we generated

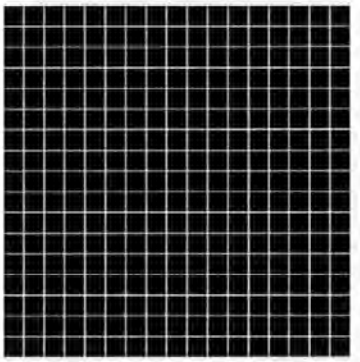
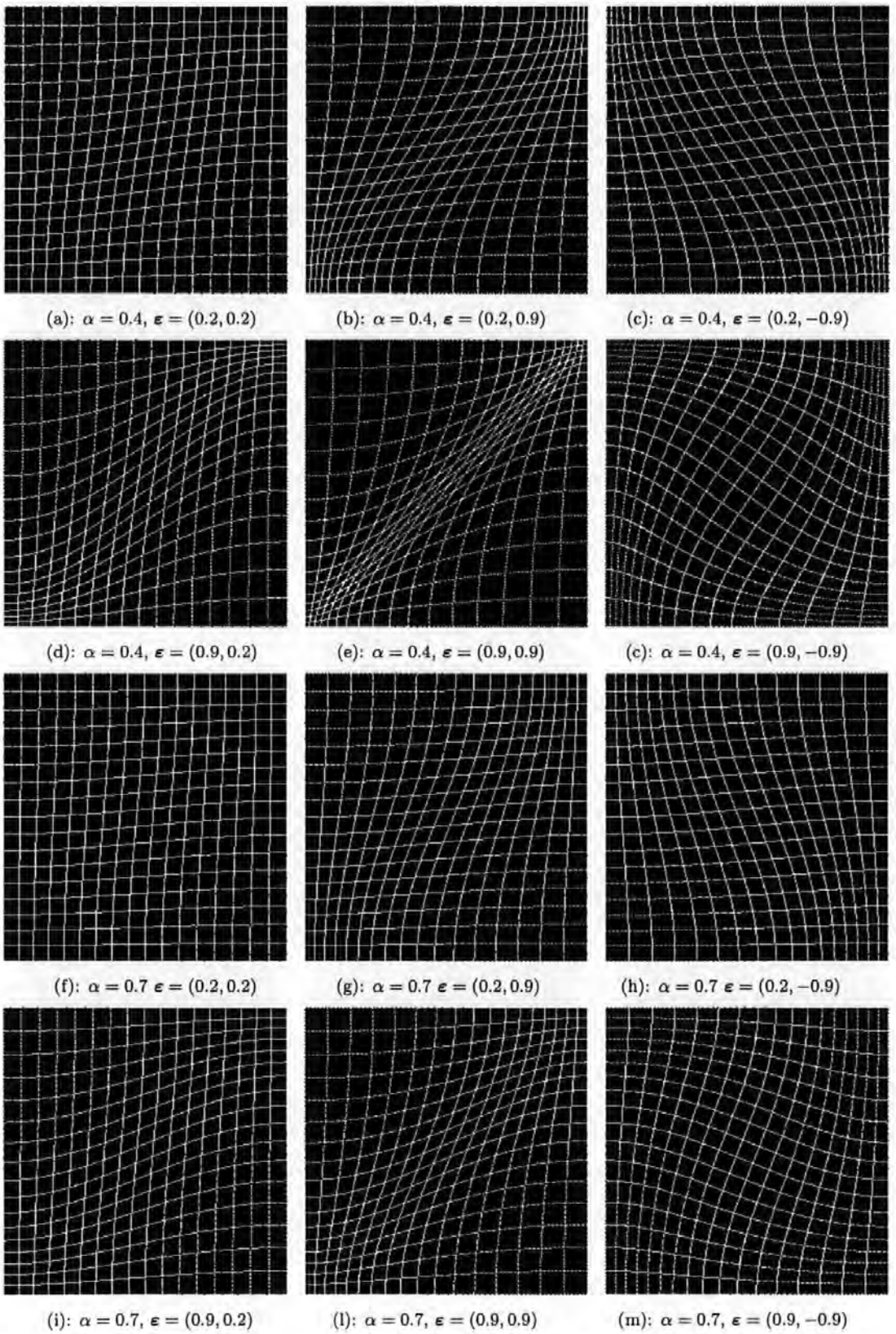


Figure 5.2: *The reference grid* $\mathbf{R}$ *drawn on* $\mathbf{S} = [0,1]^2$

a binary image $\mathbf{R}$ representing a reference grid drawn on $\mathbf{S}$, as shown in Fig. 5.2. In Fig. 5.3 we show deformations of $\mathbf{R}$ corresponding to different values of the scaling factor and the deformation parameters. The images in Fig. 5.3(a), (b) and (c) show the effect of varying ε^2 when ε^1 is a fixed small value, namely $\varepsilon^1 = 0.2$, while the images in Fig. 5.3(d), (e) and (f) are obtained using the same three values of ε^2 and a fixed large value of ε^1, namely $\varepsilon^1 = 0.9$. These six images show the deformations corresponding to a small value of α, namely $\alpha = 0.4$. For the remaining images of Fig. 5.3 we used the same values of $\boldsymbol{\varepsilon}$ but a large value of α, namely $\alpha = 0.7$. We note that for fixed α the sign and the absolute value of the deformation parameters correspond respectively to the orientation and intensity of the bending along the x axis for ε^1 and the y axis for ε^2. We can also see that the larger the value of α, the smaller is the deformation amplitude.

The pixel placed in the centre of the deformation field at the point (0.5,0.5) is not affected by any deformation as it is always mapped into itself. This means that if a pixel of the template border is at the centre during the matching algorithm, it will never move again. In order to avoid having a pixel of the shape stuck in the centre of the deformation field we decided to shift the template after every deformation by re-centring its centroid.

Now that we have specified the operator $\mathcal{K}$, we are able to describe the multi-resolution



(a): $\alpha = 0.4$, $\boldsymbol{\varepsilon} = (0.2, 0.2)$ (b): $\alpha = 0.4$, $\boldsymbol{\varepsilon} = (0.2, 0.9)$ (c): $\alpha = 0.4$, $\boldsymbol{\varepsilon} = (0.2, -0.9)$

(d): $\alpha = 0.4$, $\boldsymbol{\varepsilon} = (0.9, 0.2)$ (e): $\alpha = 0.4$, $\boldsymbol{\varepsilon} = (0.9, 0.9)$ (c): $\alpha = 0.4$, $\boldsymbol{\varepsilon} = (0.9, -0.9)$

(f): $\alpha = 0.7$ $\boldsymbol{\varepsilon} = (0.2, 0.2)$ (g): $\alpha = 0.7$ $\boldsymbol{\varepsilon} = (0.2, 0.9)$ (h): $\alpha = 0.7$ $\boldsymbol{\varepsilon} = (0.2, -0.9)$

(i): $\alpha = 0.7$, $\boldsymbol{\varepsilon} = (0.9, 0.2)$ (l): $\alpha = 0.7$, $\boldsymbol{\varepsilon} = (0.9, 0.9)$ (m): $\alpha = 0.7$, $\boldsymbol{\varepsilon} = (0.9, -0.9)$

Figure 5.3: *The effect on the reference grid* $\mathbf{R}$ *shown in Fig. 5.2 of varying the scaling factor* α *and deformation parameters* $\boldsymbol{\varepsilon}$ *of the mapping* $\mathcal{K}_{\boldsymbol{\Theta}}$ *in which* $\boldsymbol{\Theta} = (\alpha;\ \boldsymbol{\varepsilon})$.

approach for deforming the template. The basic idea is that, in order to widen the richness of the space of shapes that can be reached by the algorithm, instead of using one single deformation, we compose different transformations with different scaling factors. In particular, α increases as the algorithm proceeds. The idea is to refine the shape at successive steps using transformations that become smaller but finer. To formalize this composition of transformations, we begin by introducing the scaling parameters $\left(\alpha^{(1)}, \ldots, \alpha^{(R)}\right)$ with $\alpha^{(i)} > \alpha^{(i-1)}$, $i = 2, \ldots, R$, the set $\{\varepsilon\}$ of deformations, and the number of times $(r_1, \ldots, r_R)$ that the deformation is applied at each scale. Then our composition of transformations applied to the prototype $\mathbf{T}_0$ is given by:

$$\mathbf{T}_{com} = \prod_{i=1}^{R} \prod_{j=1}^{r_i} \mathcal{K}_{\Theta^{(i,j)}} \mathbf{T}_0, \tag{5.12}$$

where $\Theta^{(i,j)} = \left(\alpha^{(i)}; \varepsilon^{(i,j)}\right)$ in which $\varepsilon^{(i,j)}$ is a articular element of the set $\{\varepsilon\}$ of deformations. In the matching algorithm that we will describe in Section 5.5 we will choose $\varepsilon^{(i,j)}$ in an optimal way.

We now discuss 'suitable' ranges for the deformation parameters. The range $[-1, 1]$ of ε, as we said earlier, is determined by the choice of the force given in (5.9). Since we want every point of $\mathbf{S}$ to be mapped within the unit square, the minimum value of α can be obtained by requiring that the maximum deformation in both directions will never leave the unit square. We discuss this for the x direction; an analogous treatment applies for the y direction. First of all, we consider a point (x, y) very close to the border $(0, y)$. We require that this point will be mapped within $\mathbf{S}$, so that applying the maximum negative deformation in the x direction must yield a non-negative transformed point. From (5.11) and the assumption that $N = M = 1$, the maximum negative deformation corresponding to $\varepsilon^1 = -1$ must satisfy:

$$x - \frac{1}{\alpha \pi^2} \sin(\pi x) \cos(\pi y) \geq 0. \tag{5.13}$$

Since x is small we approximate $\sin(\pi x)$ by πx, and in order to have the biggest deformation we set $\cos(\pi y) = 1$. With these simplifications (5.13) becomes:

$$x - \frac{x}{\alpha\pi} = x\left(1 - \frac{1}{\alpha\pi}\right) \geq 0$$

from which it follows that $\alpha \geq \frac{1}{\pi} \sim 0.32$.

Analogously, if we take a point close to the border $(1, y)$ we require that the biggest positive deformation corresponding to $\varepsilon^1 = 1$ will not map this point beyond the border so that:

$$x + \frac{1}{\alpha\pi^2}\sin(\pi x)\cos(\pi y) \leq 1. \tag{5.14}$$

Now, since x is very close to 1, we can approximate $\sin(\pi x)$ by $\pi(1-x)$ and as before in order to have the biggest deformation we take $\cos(\pi y) = 1$. Hence, (5.14) becomes:

$$x + \frac{\pi(1-x)}{\alpha\pi^2} \leq 1,$$

leading again to the same constraint $\alpha \geq \frac{1}{\pi} \sim 0.32$. Now, because of the discretization of the mapping we will assume the value 0.4 for the minimum of α. The maximum value of α can be obtained by requiring that the maximum deformation is greater than the distance between two adjacent pixels. This happens when $\frac{1}{\alpha\pi^2} > \frac{1}{50}$ because the pixel grids that we use are typically of size 50×50. In this way, if $\alpha > \frac{50}{\pi^2} \sim 5$ every point of $\mathbf{S}$ will be mapped into itself and no effect of the deformation will be visible. Typically the size 50×50 for this kind of study is sufficiently large but if a larger support needs to be considered then the maximum value of α will be increased. In conclusion, we adopted the range $[0.4, 5]$ for the scale parameter α.

In order to illustrate the effect of composing the deformations, in Fig. 5.4 we show, as an example, the deformation obtained by setting $R = 3$, $\left(\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}\right) = (0.4, 1.5, 3)$, $r_1 = r_2 = r_3 = 1$ and adopting random choices for the deformation parameters $\varepsilon^{(i,j)}$ in

(5.12). We note that very complex deformations can be achieved by our multiresolution model. These can include approximations of rotations.

Figure 5.4: *An example of a deformation obtained as a composition of three transformations corresponding to* $(\alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}) = (0.4, 1.5, 3)$ *with random choices for the deformation parameters* $\varepsilon^{(i,j)}$.

Now that the deformation field has been introduced, we can investigate the effect of these deformations on the prototype template $\mathbf{T}_0$ shown in Fig. 5.1(b).

5.2.1 Filters to recover the connectivity

Although the operator $\mathcal{K}$ preserves the connectivity of the template on a continuous space, it may not always do so on a discrete space such as the space of image pixels. Because of the assumptions stated in Section 5.1 and the definition of the objective function that will be given in Section 5.4, to perform our analysis we need a simply connected template. In our model the range of Θ has been set in such a way that connectivity is lost for only a few extreme deformations. When this happens the connectivity may be recovered by means of standard morphological operations described in the following (see Soille, 1999). However, these standard morphological operators have the tendency to smooth the original shape, and so to overcome this drawback we developed a new filter to recover the shape connectivity. Of course, we could reduce the number of deformations generating

unconnected shapes by re-defining the image on a finer pixel grid. However, it must be noted that increasing the number of pixels to be analyzed can make the computational burden of the whole procedure too onerous. A good compromise is therefore to consider a reasonable pixel grid (in our case 50×50 pixels) and to recover template connectivity whenever is lost. To show an example where connectivity is lost, we report in Fig 5.5(a) a single deformation of $\mathbf{T}_0$ corresponding to $\Theta = (0.4; 1, -1)$. This represents one of the most extreme transformations that can be obtained by a single deformation, since α is set to its minimum value and the absolute values of the deformation parameters are at their maxima. In order to recover a simply connected shape, first we apply a morphological Bridge transformation. This transformation connects pixels previously unconnected. In Appendix B.2 we present the detail of this morphological operation and its implementation. In Fig. 5.5(b) we show the output of the morphological Bridge transformation. We note that although the shape is now connected, it is not simply connected. In fact, the Bridge filter connects the borders every time the distance between them is less than two pixels so making narrow shapes multiply connected. One way to restore simple connectivity is to fill the internal holes and then to apply the morphological Skeletonisation. We supply details about this morphological operation in Appendix B.1. The result of this filling and Skeletonisation procedure is shown in Fig. 5.5(c). This shape is now simply connected. However, if we compare this shape with the original one in Fig. 5.5(a), we note that the former is smoother than the latter. Some detailed information about the border has been lost. This is undesirable since, as we discussed in Section 1.2, the roughness of the border is an important morphological sign for classifying tumours. Therefore, in order to obtain a more satisfactory connected shape we developed a new filter. This filter recovers the external border of the shape created by the morphological Bridge. We are interested in the external border because, due to the geometry of the problem, the artifacts created

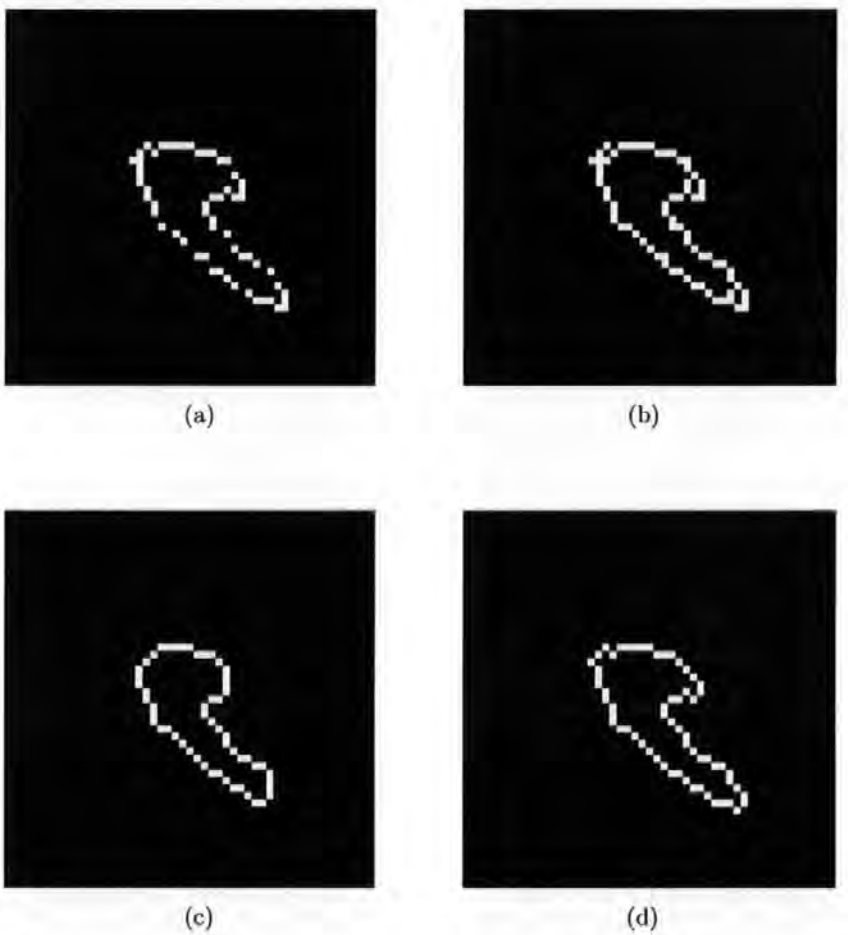


Figure 5.5: *(a) An unconnected shape obtained by applying an extreme transformation to* $\mathbf{T}_0$. *(b) The output of the morphological Bridge filter. The shape is now connected but not simply connected. (c) The output of the morphological hole-filling and Skeletonisation filters. This shape is smoother than the one shown in (a). Some information about the contour detail has been lost. (d) The result of applying the new Ext filter. The shape is simply connected and resembles more the one shown in (a). No information about the contour roughness has been lost.*

by this morphological transformation are very likely to be in the interior of the template. Our new filter can be described as follows:

1. following a fixed sweeping scheme, all the image pixels are visited until a shape pixel is found. This starting shape pixel becomes the current pixel and forms the first pixel of the filtered shape.

2. The second order neighbours of the current pixel are considered following a particular visiting scheme. The first shape pixel that has not yet been visited is selected. The current pixel becomes the old pixel and the selected pixel becomes the current pixel. The current pixel is added to the pixel of the filtered shape. The visiting scheme depends on the curvature of the shape at the current pixel and this is estimated by the relative location of the current and old pixel. Some visiting schemes are reported in Fig. 5.6. For the starting pixel of Step 1 a fixed visiting scheme is assumed.

3. Step 2 is repeated until the current pixel is a neighbour of the starting pixel. This always happens since the shape obtained by applying the Bridge filter is connected and the direction (in our case anticlockwise) is maintained throughout the algorithm.

1	old	7
2	cur	6
3	4	5

old	7	6
1	cur	5
2	3	4

3	2	1
4	cur	old
5	6	7

Figure 5.6: *Visiting scheme for neighbouring pixels. The visiting order depends on the local curvature, estimated by the relative position of the current pixel (cur) and the old pixel (old). Three examples are shown in this figure. The first visiting scheme is the one adopted for the starting pixel, except old would be replaced by 8.*

In the following we indicate this new filter as the Ext filter. The result of this filter will always be the same when it is applied after the Bridge filter, provided that the starting shape pixel is one of the pixels of the outer border. In Fig. 5.5(d) we show the shape recovered by this filter. We note that our Ext filter has preserved the features of the border of the original shape shown in Fig. 5.5(a) much better than the previous morphology based approach shown in Fig. 5.5(c). The Ext filter is also as fast as the morphological operations (typically 1.6×10^{-2} sec per reconstruction), and is implemented in one step instead of two.

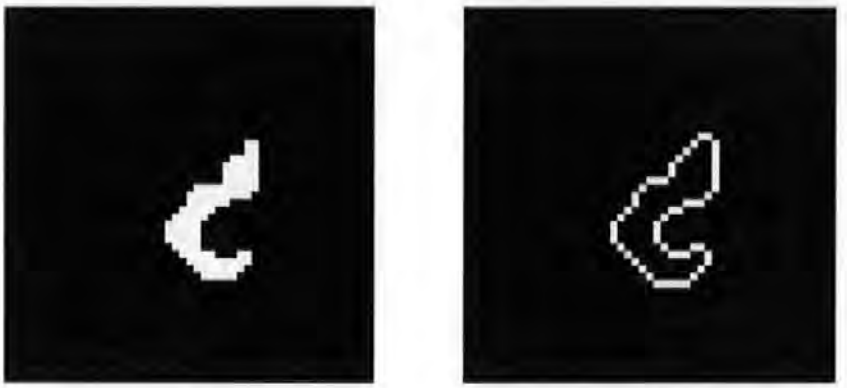
5.3 The richness of the deformation model

In this section we present a simulation study to investigate the richness of the space of shapes generated by the deformation model (5.12). We investigate how versatile our deformation model is in reaching a particular shape (that we call the 'true' shape in the following) obtained independently of the deformation model itself. In addition, we want to compare the general model that uses a composition of deformations with the one that uses a single deformation ($R = 1$, $r_1 = 1$). In order to test these two models on a reasonable approximation of the real lesion shape, we decided to extract the lesion contour from the classification image shown in Fig. 5.7. This image represents the classification result within $\mathcal{I}$ obtained with the non-parametric Bayesian classification using only the attributes $\mathbf{A} = (A_i)$. These attributes contain information about the contrast agent wash-in that is one of the main features for discriminating between malignant and benign tumoral tissues. At this stage of our study the matching algorithm is based on one classification attribute but in the future we will extend this approach to cope with more attributes. In Fig. 5.7 the green region represents the pixels classified as malignant tumoral tissue. As we said earlier, in the following we will refer to this region as the lesion. In order to extract from

Figure 5.7: *Non-parametric Bayesian classification of Chapter 4 obtained using only the attributes* **A** *within the tumoral region* $\mathcal{I}$ *superimposed on the last image of the sequence. The green and red regions represent the malignant and benign tumoral tissues respectively.*

this image a simply connected shape approximating the lesion border, we first applied a morphological Closure (see Soille, 1998) to the green region of this figure; the result is shown in Fig. 5.8(a). We then extracted the border of this region. This contour, shown in Fig. 5.8(b), is assumed to represent the 'true' shape to be reconstructed.

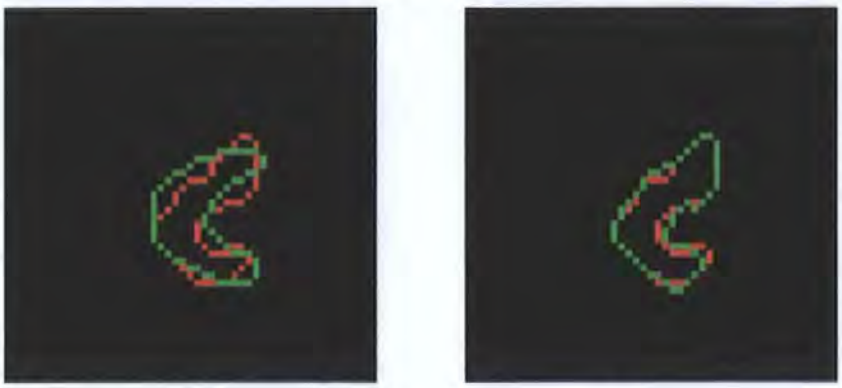
As we have mentioned earlier, in our procedure the deformation is driven by the minimization over the deformation parameters of a particular objective function. In this section we want to investigate the richness of the deformation model and we do not want this study to be influenced by the effect of the particular choice made for the objective function. In this case, since the true shape is known, we adopt an objective function that depends on the quality of the matching between the 'true' and reconstructed shape. In particular, the objective function will be the number of mismatched pixels between the interior of the 'true' and the reconstructed shape. With this choice of objective function any difference between the 'true' and reconstructed shape will depend exclusively on the inability of the model to generate the 'true' shape starting from $\mathbf{T}_0$. In Fig.5.9(a) we show



(a) (b)

Figure 5.8: *(a) The malignant tumoral region after the morphological Closure operation. (b) The border of the malignant tumoral region that will be taken as the 'true' shape to be reconstructed in this simulation.*

the reconstruction obtained by using a single deformation model. The final shape is closer to the 'true' shape than $\mathbf{T}_0$, but we note a number of mismatched pixels. In particular, we obtained this result using the minimum of α, namely $\alpha = 0.4$. This value corresponds to the largest deformation amplitude. We also explored the whole set $\{\ \varepsilon\}$. The value of the objective function for this reconstruction was 76 mismatched pixels. Since the size of the interior of the 'true' shape is 110 pixels, we obtained a result of 70% mismatched pixels. If instead we apply a composition of deformation, we may obtain a more accurate reconstruction such as the one shown in Fig. 5.9(b). For this reconstruction we set $R = 47$ and $\left(\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(47)}\right) = (0.4, 0.5, \ldots, 5)$ and at each scale the transformation was applied until there were no further changes. The agreement between the reconstructed and 'true' shape is now very good as the finer deformation scales have refined the final shape. The value of the objective function has reduced to 23 (20%) mismatched pixels. We therefore use a composite deformation of exactly this type in the optimization algorithm described in Section 5.5.



(a) (b)

Figure 5.9: *(a) Comparison between the 'true' shape (red) and its reconstruction (green) obtained using a single deformation model. (b) Comparison between between the 'true' shape (red) and its reconstruction (green) obtained using a composition of deformations.*

5.4 The objective function

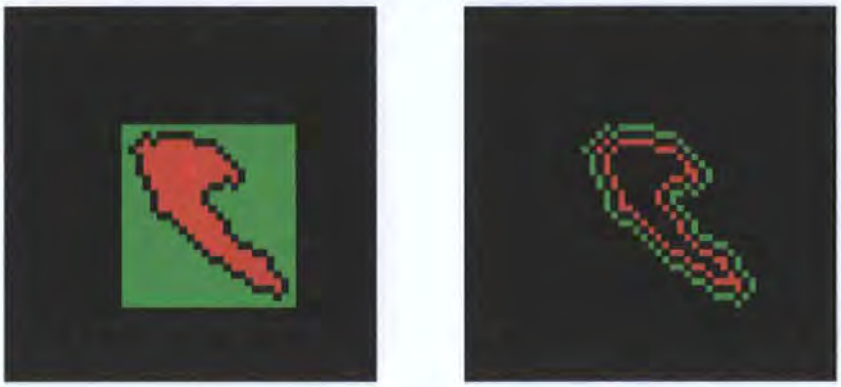
In this section we introduce the objective function that drives the template deformation in the matching algorithm. In fact, in order to reconstruct the lesion contour, this function is minimized over the deformation parameters.

As in the Bayesian classification approach, our matching procedure is based on the classification attributes. In particular, in this study we adopted the attributes **A** described in Chapter 4. Among the different choices of the objective function that we investigated, the one that performed best is given by:

$$V = \beta V_{edges} + V_{test}, \tag{5.15}$$

where V_{edges} is a penalty based on the edge of the current shape, V_{test} is a penalty based on a local hypothesis test and $\beta > 0$ is a scaling parameter. We now describe V_{edges} and V_{test} in detail. In order to do this we need to introduce some notation. First, for a given shape $\mathbf{T}$ we define the interior $\mathtt{I}(\mathbf{T})$ to be the set of pixels inside it. These pixels are coloured in red in Fig. 5.10(a). On the other hand, the exterior $\mathtt{E}(\mathbf{T})$ is defined to be all the pixels

in the smallest rectangular region containing $\mathbf{T}$ excluding the pixels that comprise the contour of $\mathbf{T}$ and $\mathtt{I}(\mathbf{T})$. These pixels are shown in green in Fig. 5.10(a). Given $\mathtt{I}(\mathbf{T})$ and



(a) (b)

Figure 5.10: *(a) The interior $\mathtt{I}(\mathbf{T})$ (red) and the exterior $\mathtt{E}(\mathbf{T})$ (green) of the shape $\mathbf{T}$ shown in black. (b) The internal edge $\mathbf{edge}(\mathtt{I}(\mathbf{T}))$ (red) and the external edge $\mathbf{edge}(\mathtt{E}(\mathbf{T}))$ (green).*

$\mathtt{E}(\mathbf{T})$, we define the internal edge $\mathbf{edge}(\mathtt{I}(\mathbf{T}))$ and the external edge $\mathbf{edge}(\mathtt{E}(\mathbf{T}))$ to be the second order neighbours of the contour of $\mathbf{T}$ that belong to $\mathtt{I}(\mathbf{T})$ and $\mathtt{E}(\mathbf{T})$ respectively. The pixel sets $\mathbf{edge}(\mathtt{I}(\mathbf{T}))$ and $\mathbf{edge}(\mathtt{E}(\mathbf{T}))$ are shown in red and green respectively in Fig. 5.10(b). We also define $\mu_{\mathtt{I}(\mathbf{T})}$, $\mu_{\mathtt{E}(\mathbf{T})}$, $\sigma_{\mathtt{I}(\mathbf{T})}$ and $\sigma_{\mathtt{E}(\mathbf{T})}$ to be the means and the standard deviations of the classification attributes $\mathbf{A}$ in $\mathtt{I}(\mathbf{T})$ and $\mathtt{E}(\mathbf{T})$. We are now able to define the edge penalty $V_{edges}(\mathbf{T})$ as:

$$V_{edges}(\mathbf{T}) = \frac{\sum_{i \in \mathbf{edge}(\mathtt{I}(\mathbf{T}))} \left\{ A_i - \mu_{\mathtt{I}(\mathbf{T})} \right\}^2}{2\sigma^2_{\mathtt{I}(\mathbf{T})}} + \frac{\sum_{i \in \mathbf{edge}(\mathtt{E}(\mathbf{T}))} \left\{ A_i - \mu_{\mathtt{E}(\mathbf{T})} \right\}^2}{2\sigma^2_{\mathtt{E}(\mathbf{T})}}. \tag{5.16}$$

Note that (5.16) is based on a Gaussian model for the distributions of the classification attributes $\mathbf{A}$ inside and outside the lesion $\mathbf{T}$. As our algorithm proceeds, $\mathbf{T}$ varies with the results that the parameters $\mu_{\mathtt{I}(\mathbf{T})}$, $\mu_{\mathtt{E}(\mathbf{T})}$, $\sigma_{\mathtt{I}(\mathbf{T})}$ and $\sigma_{\mathtt{E}(\mathbf{T})}$ of these distributions are updated.

The rationale behind (5.16) is that the sum of the squared differences between the attribute values at internal (external) edge pixels and $\mu_{\mathtt{I}(\mathbf{T})}$ ($\mu_{\mathtt{E}(\mathbf{T})}$) should be minimized

when the lesion contour is estimated correctly. We can think of the lesion contour as being driven by its internal and external edges, with both these edges playing an important role. In fact, if we use only the first term in (5.16) we may obtain an estimate of the contour that is completely inside the lesion. We may also be unable to discriminate between different contours inside the lesion which have comparably low values of the objective function. However, if we also include the second term in (5.16) such problems should not arise because for such contours this term is large. Therefore, the external edge 'pulls' the contour towards the outside. Analogous considerations apply if we were to use only the second term in (5.16) when defining $V_{edges}(\mathbf{T})$. In the penalty model (5.16) the summation is over $\mathtt{edge}(\mathtt{I}(\mathbf{T}))$ and $\mathtt{edge}(\mathtt{E}(\mathbf{T}))$ instead of $\mathtt{I}(\mathbf{T})$ and $\mathtt{E}(\mathbf{T})$ because in this way $V_{edges}(\mathbf{T})$ is more sensitive to small changes in the lesion contour. However, we base the calculation of the means and standard deviations on the whole of the interior $\mathtt{I}(\mathbf{T})$ and the exterior $\mathtt{E}(\mathbf{T})$. This causes the estimate of these parameters and hence the matching algorithm itself to be more robust to outliers in the random noise.

We now describe in detail how we estimate the interior region $\mathtt{I}(\mathbf{T})$. In fact, this task can be very difficult when dealing with shapes $\mathbf{T}$ the contours of which are very rough such as some of the shapes that can be generated by the deformation model (5.12). For this reason, we developed a new method based essentially on the idea behind the Ext filter presented in Section 5.2.1. The idea is that to recover successfully the internal region of a shape we must visit the shape pixels in a fixed order that takes into account the local curvature of the shape. As in the case of the Ext filter, the algorithm proceeds in an anticlockwise orientation estimating the local curvature from the coordinates of the current and old shape pixel. The method works as follows. Every shape pixel is considered according to the visiting scheme presented in Fig. 5.6. From pixel (i, j) the algorithm moves along the i^{th} row, going to the left or to the right depending on the local

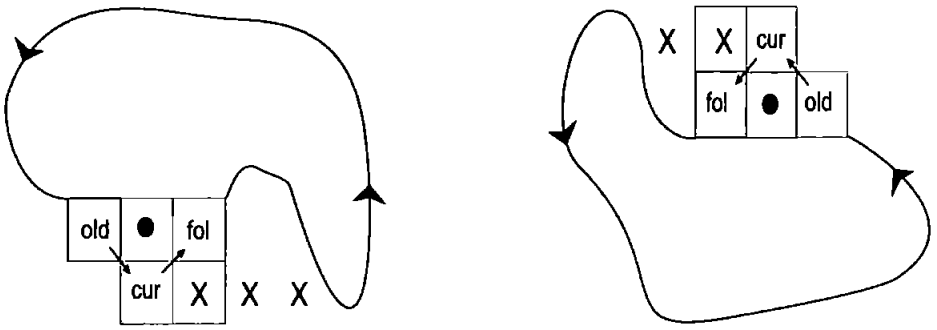


Figure 5.11: *Two cases in which the orientation and the curvature estimated from the current (cur) and old (old) pixels are not sufficient to find the internal pixels of the shape. The pixels that are wrongly identified as internal are indicated by an X. The internal points found using the current, old and following pixel (fol) are indicated by a black dot.*

curvature, and includes in $\mathtt{I}(\mathbf{T})$ every pixel until another shape pixel (i, j^*) is met. In other words, all the pixels (i, j') with $j^* < j' < j$ or $j < j' < j^*$ are included in $\mathtt{I}(\mathbf{T})$. If in the i^{th} row there is just one shape pixel, no pixel is added to $\mathtt{I}(\mathbf{T})$. This scheme identifies the internal pixels correctly except in the case of the two neighbourhood configurations shown in Fig. 5.11. As we can see from this figure, the orientation and the estimated local curvature are not sufficient for identifying the interior in these two configurations. In fact, without modification the above scheme would add pixels to $\mathtt{I}(\mathbf{T})$, labelled X in the figure, that are outside the shape. In these cases, to identify the internal pixels correctly we need also to take into account the 'following' (labelled fol) shape pixel, and not just the current and previous ones (labelled cur and old respectively). In this way, considering the configuration of the previous, current and following pixels, would lead us correctly to include in $\mathtt{I}(\mathbf{T})$ only the pixel above or below the current one (indicated by a black spot).

We now describe the second penalty term V_{test} introduced in (5.15). Since the tumour is characterized by two homogenous regions separated by a discontinuity line that identifies the lesion contour, the idea is to attract the contour estimate towards this discontinuity by means of an hypothesis test that checks whether each pixel of the current shape contour

marks the discontinuity. For each shape pixel i the test is conducted using the values of the attribute image $\mathbf{A}$ at the pixels $\mathcal{N}(i, \mathrm{I})$ and $\mathcal{N}(i, \mathrm{E})$ identified by the internal and external unit normal vectors centred at i. Then V_{test} takes into account the number of shape pixels $n(\mathbf{T})$ that are identified by the test as marking the discontinuity in this way:

$$V_{test} = 1 - \frac{n(\mathbf{T})}{N(\mathbf{T})},$$

where $N(\mathbf{T})$ is the total number of pixels that comprise the shape contour. Let us now define the test procedure. We test the null hypothesis H_0 that between $\mathcal{N}(i, \mathrm{I})$ and $\mathcal{N}(i, \mathrm{E})$ there is no discontinuity (i.e these pixels belong to the same region) against the alternative hypothesis H_1 that between $\mathcal{N}(i, \mathrm{I})$ and $\mathcal{N}(i, \mathrm{E})$ there is a discontinuity (i.e. these pixels do not belong to the same region). An assumption behind our test is that $\mu_{\mathrm{I}(\mathbf{T})}$, $\mu_{\mathrm{E}(\mathbf{T})}$, $\sigma_{\mathrm{I}(\mathbf{T})}$ and $\sigma_{\mathrm{E}(\mathbf{T})}$ represent good estimates of the parameters of the distributions of $\mathbf{A}$ within I and E. For notational simplicity, we now suppress the dependence on $\mathbf{T}$. If there is no edge between $\mathcal{N}(i, \mathrm{I})$ and $\mathcal{N}(i, \mathrm{E})$ we assume that either $A_{\mathcal{N}(i,\mathrm{I})} - A_{\mathcal{N}(i,\mathrm{E})} \sim \mathrm{N}(0, 2\sigma_{\mathrm{I}}^2)$ or $A_{\mathcal{N}(i,\mathrm{I})} - A_{\mathcal{N}(i,\mathrm{E})} \sim \mathrm{N}(0, 2\sigma_{\mathrm{E}}^2)$. If instead there is an edge, we can assume that $A_{\mathcal{N}(i,\mathrm{I})} - A_{\mathcal{N}(i,\mathrm{E})} \sim \mathrm{N}(\mu_{\mathrm{I}} - \mu_{\mathrm{E}}, \sigma_{\mathrm{I}}^2 + \sigma_{\mathrm{E}}^2)$.

Thus, to test these hypotheses we adopted the following test statistic:

$$W = \frac{A_{\mathcal{N}(i,\mathrm{I})} - A_{\mathcal{N}(i,\mathrm{E})}}{\sqrt{\sigma_{\mathrm{I}}^2 + \sigma_{\mathrm{E}}^2}}. \tag{5.17}$$

Since we assume for the simulated data that the values of the classification attribute are higher inside the lesion than outside, we have that $A_{\mathcal{N}(i,\mathrm{I})} \geq A_{\mathcal{N}(i,\mathrm{E})}$ and so we adopt a one-tailed test: reject H_0 if $W \geq w$. We choose w by considering the power of the test:

$$\mathrm{P}(\mathrm{H}_0 \text{ is rejected} | \mathrm{H}_1 \text{ is true}) = \mathrm{P}\left(W \geq w | W \sim \mathrm{N}\left(\frac{\mu_{\mathrm{I}} - \mu_{\mathrm{E}}}{\sqrt{\sigma_{\mathrm{I}}^2 + \sigma_{\mathrm{E}}^2}}, 1\right)\right).$$

In particular, let us require power λ when $\dfrac{\mu_{\mathrm{I}} - \mu_{\mathrm{E}}}{\sqrt{\sigma_{\mathrm{I}}^2 + \sigma_{\mathrm{E}}^2}} = \delta > 0$. That is, we require $\mathrm{P}(W \geq w | W \sim \mathrm{N}(\delta, 1)) = \lambda$. Hence, $\mathrm{P}(W - \delta \geq w - \delta | W - \delta \sim \mathrm{N}(0, 1)) = \lambda$, so that

$1 - \Phi(w - \delta) = \lambda$, or $w = \delta + \Phi^{-1}(1 - \lambda)$, where Φ is the cumulative distribution function of a standard normal variable.

An upper bound on the probability of Type I error can be found as follows:

$$\mathrm{P}(\mathrm{H_0} \text{ is rejected} | \mathrm{H_0} \text{ is true}) = 1 - \frac{1}{2}\left\{\Phi\left(\frac{w}{\sqrt{2}}\sqrt{1 + \frac{\sigma_{\mathrm{E}}^2}{\sigma_{\mathrm{I}}^2}}\right) + \Phi\left(\frac{w}{\sqrt{2}}\sqrt{1 + \frac{\sigma_{\mathrm{I}}^2}{\sigma_{\mathrm{E}}^2}}\right)\right\},$$

assuming that under $\mathrm{H_0}$ the probability that both pixels belong to I (or E) is $\frac{1}{2}$. Elementary calculus leads us to the bound:

$$\mathrm{P}(\mathrm{H_0} \text{ is rejected} | \mathrm{H_0} \text{ is true}) \leq 1 - \Phi(w),$$

with equality holding when $\sigma_{\mathrm{I}} = \sigma_{\mathrm{E}}$. Hence, an upper bound for the probability of Type I error is $1 - \Phi\{\delta + \Phi^{-1}(1 - \lambda)\}$. In our simulation study a conservative value of δ is 1.5. Thus, for power $\lambda = 0.7$, we can set $w = 1$. This choice leads to an upper bound for the probability of Type I error of 0.16, which is acceptable. With real data we do not know 'a priori' that values of the classification attribute are higher inside the lesion than outside. Accordingly, with real data we must modify (5.17) taking into account the absolute value of the difference between the attribute values.

We remark that V_{test} takes values in $[0, 1]$. In particular, the larger the number of shape pixels identified by the test as marking the discontinuity, the smaller is the value of V_{test}, with zero being reached when $n(\mathbf{T}) = N(\mathbf{T})$. Since V_{edges} is not guaranteed to lie in $[0, 1]$ we may need to scale it so that it will assume comparable values to V_{test} in the objective function V. To do this, for every dataset we first run the matching procedure with $V = V_{edges}$. Then, by using the maximum value M of V_{edges} obtained during this initial phase, we scale V_{edges} by setting $\beta = \frac{1}{M}$ in (5.15). We also note here that the whole objective function (5.15) is bounded below by zero.

As mentioned above, the objective function V defined in (5.15) gave the best quality

lesion estimates. We experimented using just one of the terms V_{edges} and V_{test}, but in both cases inferior estimates were obtained.

5.5 The matching algorithm

In this section we present details of the matching algorithm. In particular, we describe the initial version of the algorithm and the final modification of it that improved the performance in terms of stopping time and quality of the lesion estimate.

As explained in Section 5.2, the minimum and maximum possible values of α are set to 0.4 and 5 respectively. We decided to adopt the whole range $[0.4, 5]$ for α and set $R = 47$ yielding the parameter sequence $\left(\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(R)}\right) = (0.4, 0.5, \ldots, 5)$. At each iteration j of scale $\alpha^{(i)}$, the current shape is deformed over the set of deformation parameters $\{\varepsilon\}$ and the deformation parameters corresponding to the deformed shape that minimizes the objective function are called $\varepsilon^{(i,j)}$. If $\varepsilon^{(i,j)}$ is different from zero the whole process is repeated. The number of iterations that result in a deformed shape that corresponds to a lower minimum of the objective function is denoted by r_i. We note that $r_i < \infty$ because the number of possible new shapes that can be generated at each scale is finite, there being a limited number of possible simply connected shapes defined on our pixel grid.

When no further reduction of the objective function occurs, we increase i so changing the scale to $\alpha^{(i+1)}$ and repeat the above process until all the scales have been considered. It turns out that often the algorithm stopped (in the sense that no further reduction of the objective function is achieved) before the last value of the scale parameter.

The matching algorithm is represented schematically by the flow chart shown in Fig. 5.12. Whenever a shape is deformed, its connectivity is checked. If the deformed shape is not simply connected the Bridge filter and the Ext filter are applied in sequence as described in Section 5.2.1.

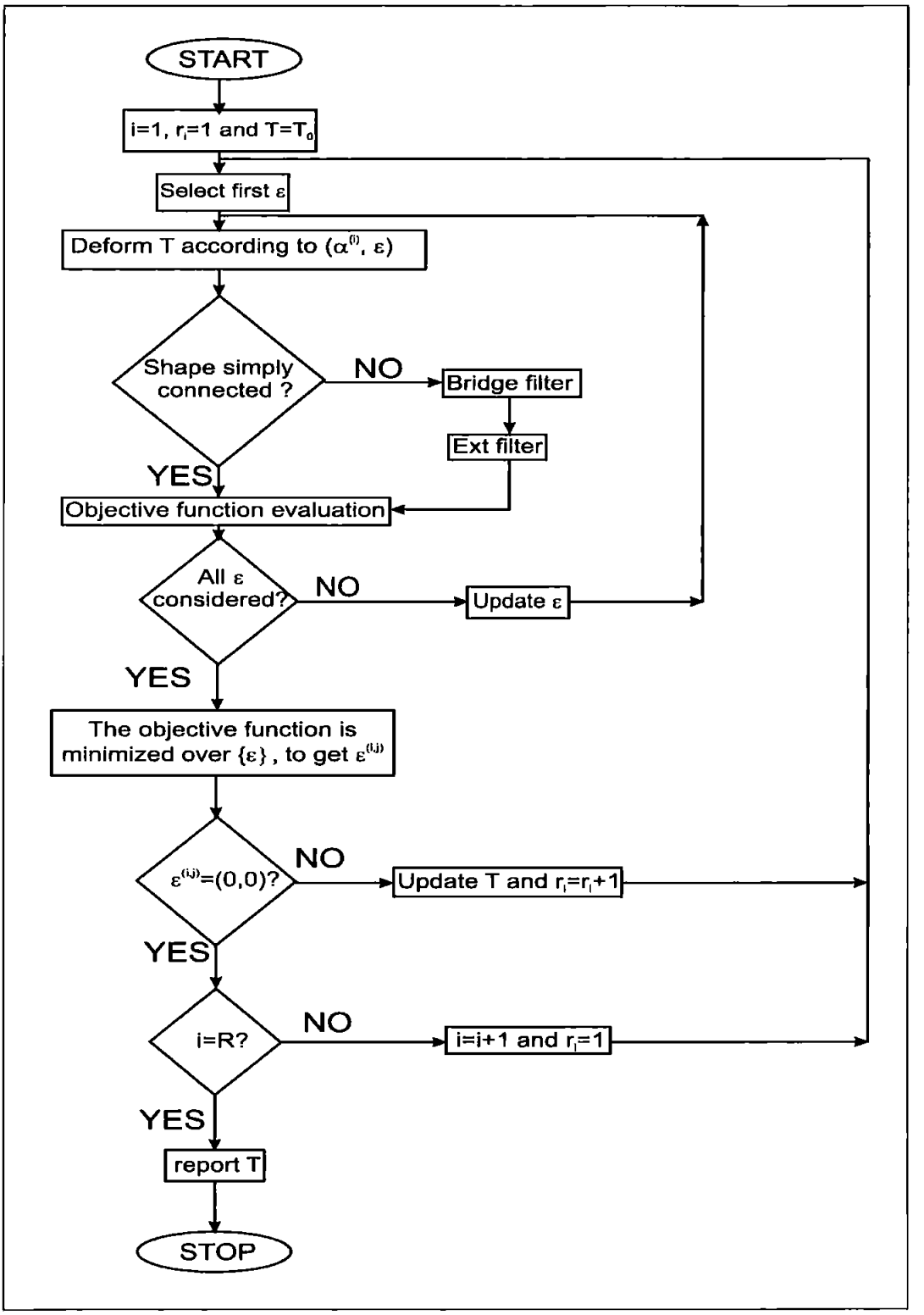


Figure 5.12: *A flowchart representing the matching algorithm.*

As will be described in the next section, this version of the algorithm showed some limitations. In fact, this version seems not able to perform local deformations of the shape in the sense that it cannot transform just some parts of the shape, leaving the rest unchanged. Often a parameter choice leading to a shape reconstruction that was optimal apart from a few mismatched pixels far from the lesion was refused in favour of a worse reconstruction with more mismatched pixels closer to the lesion. In order to overcome this limitation, we decided to modify the algorithm as follows: at the end of every iteration the hypothesis test described in Section 5.4 is performed on the updated current shape. Then, shape pixels that are identified by the test as marking the discontinuity are 'locked', i.e. these pixels cannot be moved during the next iteration. In this way, deformations at the next step will be applied only to the remaining 'unlocked' pixels of the shape. Since the test is repeated at the end of every iteration and is based on both the current optimal shape and corresponding estimates of μ_{I}, μ_{E}, σ_{I} and σ_{E}, locked pixels at a given step can be unlocked at the following step. In fact, if at the end of an iteration no pixels are identified as marking the discontinuity, none will be locked at the following step and this part of the procedure remains the same as before. The advantage of the modified algorithm is that it can now perform both local and global deformations whilst the original version was restricted to global deformations. Furthermore, as we will illustrate in the following section, the modified version of the algorithm can sometimes produce reliable results much faster.

5.6 Results

In this section we present results obtained using both synthetic and real data. In the first case, we simulate the classification attribute image $\mathbf{A}$ in order to investigate the potential of the objective function with a known true lesion. In the second case, we use $\mathbf{A}$ obtained

from the non-parametric Bayesian classification described in Section 4.

5.6.1 Synthetic data

In this study we produced two types of simulated data. Both of them are obtained by filling the interior and the exterior of a given 'true' shape with values sampled from two Gaussian distributions with known parameters μ_{I}, μ_{E}, σ_{I} and σ_{E}. By varying these parameters we can investigate the performance of our matching algorithm on data characterized by different values of the Signal to Noise Ratio (SNR)[1].

The difference between the two types of data lies in the choice of the true shape. In the first type, the true shape is obtained by applying the deformation model (5.12) to the prototype template $\mathbf{T}_0$ shown in Fig. 5.1. In this way, the true shape is generated by deforming the contour that is the initial shape for the matching algorithm. This step is performed by using the same deformations employed in the algorithm. This largely reduces the difficulty of the reconstruction task. However, the second type of synthetic data is characterized instead by a shape that has been obtained independently from the deformation model (5.12) and the prototype template $\mathbf{T}_0$. In this case the reconstruction task is more difficult. In both cases the initial shape of the matching algorithm is $\mathbf{T}_0$.

Case 1

In Fig. 5.13 we show the true shape obtained from $\mathbf{T}_0$ after applying a deformation with parameters $R = 2$, $r_1 = r_2 = 1$, $\alpha^{(1)} = 0.7$, $\boldsymbol{\varepsilon}^{(1,1)} = (-0.9, 0.4)$, $\alpha^{(2)} = 0.9$ and $\boldsymbol{\varepsilon}^{(2,1)} = (-0.9, 0.4)$. In this case the data are generated by filling the interior and exterior regions I and E with values drawn from $\mathrm{N}\left(6, (0.5)^2\right)$ and $\mathrm{N}\left(2, (0.5)^2\right)$ distributions

[1]The SNR is a measure of signal strength relative to background noise. The ratio is usually measured in decibels (dB). If the signal strength is V_s, and the noise level is V_n, then the SNR in decibels is given by $\mathrm{SNR} = 20\log(V_s/V_n)$. In this study we considered as signal strength V_s the difference between the means of **A** inside and outside the lesion $\mu_{\mathrm{I}} - \mu_{\mathrm{E}}$ and as noise level V_n the standard deviation of the noise distribution.

Figure 5.13: *True shape obtained by applying deformation (5.12) to the prototype template* $\mathbf{T}_0$.

respectively. The resulting data are shown in Fig. 5.14(a). The background region has been selected in such a way that it is typically large enough to contain all the shapes that will be generated during the algorithm. As we can see from the high contrast of the image, this dataset is characterized by a high SNR, namely $20\log(4/0.5) \simeq 18$ dB, and so reconstructing the lesion contour should be an easy task. In fact, as we can see in Fig 5.14(b) in which the reconstructed (green) and the true (red) lesion contour are shown, we obtained an excellent agreement between them.

(a) (b)

Figure 5.14: *(a) Data generated from the true shape shown in Fig. 5.13 by filling its interior and exterior with values drawn from* $\mathrm{N}\left(6,(0.5)^2\right)$ *and* $\mathrm{N}\left(2,(0.5)^2\right)$ *distributions respectively. (b) The reconstructed (green) and true (red) lesion contour.*

Case 1			Estimated value based on the true shape	Estimated value based on reconstruction
SNR=18 dB	Region I	μ_I	5.97	5.91
		σ_I	0.51	0.63
	Region E	μ_E	2.00	2.03
		σ_E	0.49	0.64

Table 5.1: *Estimates of the distribution parameters based on the true and reconstructed shape from the data shown in Fig. 5.14(a).*

We need to specify that in all the datasets simulated in this study, the shape pixels themselves are assigned values sampled from a $N(\mu_E, \sigma_E^2)$ distribution, i.e. shape pixels are assumed to belong to the exterior of the shape. As a consequence of this, the objective function is not able to discriminate between the true lesion contour and its internal edge. This is due to the fact that in theory the contour should be infinitely narrow but in reality it has a finite size and its pixels belong either to the internal or external region. We made the arbitrary choice to assign edge pixels to the external region. Therefore, green pixels in Fig 5.14(a) that belong to the internal edge of the true contour should not be considered as misclassified. This leads to a percentage of 99% of contour pixels estimated correctly. In Table 5.1 we report the estimates of μ_I, μ_E, σ_I and σ_E based on the true and reconstructed shape. We note a very good agreement between them.

We tested the performance of the algorithm with many different noise realizations generated in the same way. We always obtained very good reconstructions comparable to the one shown.

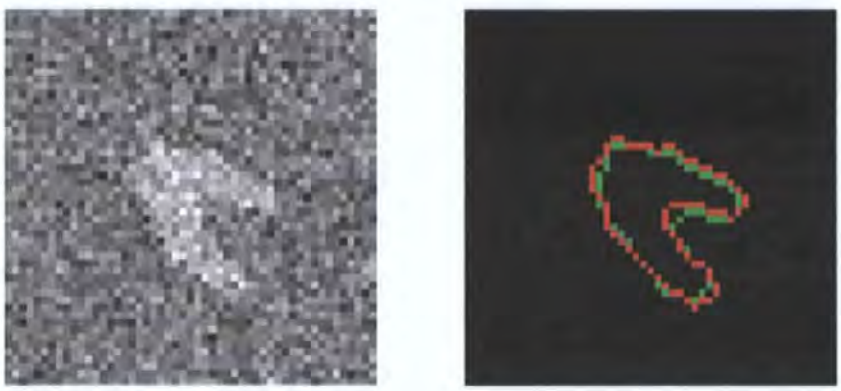
Case 2

We now investigate the performance of the algorithm with a more difficult task. We use the same true shape shown in Fig. 5.13 but generate data in such a way that the SNR is three times smaller. In particular, the data are generated by sampling from

Case 2			Estimated value based on the true shape	Estimated value based on reconstruction
SNR=6 dB	Region I	μ_I	2.98	2.91
		σ_I	0.52	0.61
	Region E	μ_E	1.99	2.02
		σ_E	0.49	0.51

Table 5.2: *Estimates of the distribution parameters based on the true and reconstructed shape from the data shown in Fig. 5.15(a).*

$N(3,(0.5)^2)$ and $N(2,(0.5)^2)$ distributions for the interior and exterior of the true shape so that SNR $= 20\log(1/0.5) = 6$ dB. In Fig 5.15(a) the simulated dataset is shown. Despite the low value of the SNR, we note from Fig. 5.15(a) a good agreement between the reconstructed (green) and true (red) lesion contour. In this case 81% of the contour



(a) (b)

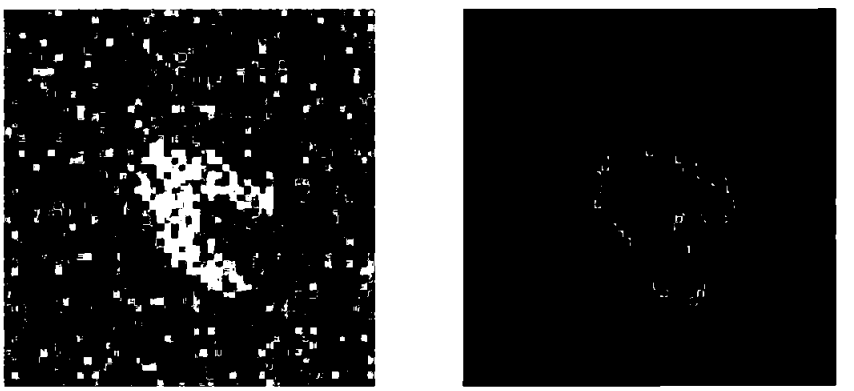
Figure 5.15: *(a) Data generated from the true shape shown in Fig. 5.13 by filling its interior and exterior with values drawn from* $N(3,(0.5)^2)$ *and* $N(2,(0.5)^2)$ *distributions respectively. (b) The reconstructed (green) and true (red) lesion contour.*

pixels are estimated correctly. Furthermore, if we compare the estimates for the parameters based on the reconstructed and true shape, reported in Table 5.2, we note that even if the SNR is very low we still obtain a very good agreement between them. Again different noise realizations lead to similar results.

Case 2			Estimated value based on the true shape	Estimated value based on reconstruction
SNR=6 dB	Region I	μ_I	2.98	2.91
		σ_I	0.52	0.61
	Region E	μ_E	1.99	2.02
		σ_E	0.49	0.51

Table 5.2: *Estimates of the distribution parameters based on the true and reconstructed shape from the data shown in Fig. 5.15(a).*

$N(3,(0.5)^2)$ and $N(2,(0.5)^2)$ distributions for the interior and exterior of the true shape so that SNR $= 20\log(1/0.5) = 6$ dB. In Fig 5.15(a) the simulated dataset is shown. Despite the low value of the SNR, we note from Fig. 5.15(a) a good agreement between the reconstructed (green) and true (red) lesion contour. In this case 81% of the contour



(a) (b)

Figure 5.15: *(a) Data generated from the true shape shown in Fig. 5.13 by filling its interior and exterior with values drawn from* $N\left(3,(0.5)^2\right)$ *and* $N\left(2,(0.5)^2\right)$ *distributions respectively. (b) The reconstructed (green) and true (red) lesion contour.*

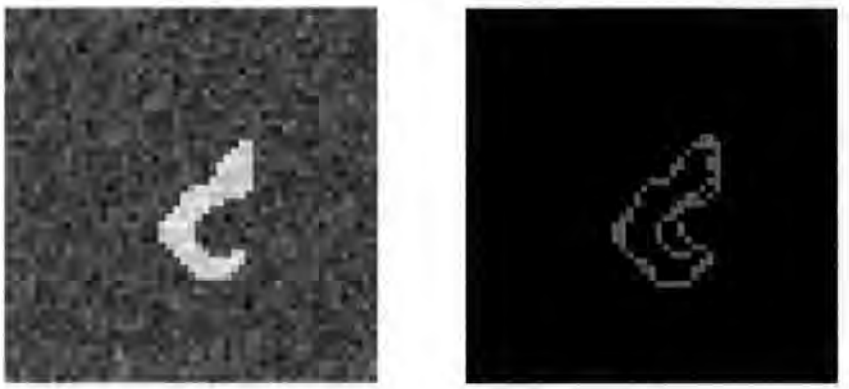
pixels are estimated correctly. Furthermore, if we compare the estimates for the parameters based on the reconstructed and true shape, reported in Table 5.2, we note that even if the SNR is very low we still obtain a very good agreement between them. Again different noise realizations lead to similar results.

Case 3

We now show the results obtained with the second type of simulated data. Here the true shape is the one introduced in Section 5.3 and shown again in Fig. 5.16. As before we generated a dataset by filling the interior and the exterior with values drawn from a Gaussian distribution with the same parameters that we used in Case 1. In this way, we can compare the results with those obtained before. In Fig. 5.17 we show data with

Figure 5.16: *The second true shape. This shape is introduced in Section 5.3.*

SNR = 18 dB and the reconstructed (green) and true (red) lesion contours. In this case 89% of contour pixels are still identified correctly despite the low SNR. Unlike with the first shape, some of the misclassified pixels are not very close to the true contour. This shows a limitation of the algorithm: it is not able to modify the shape contour locally without increasing the value of the objective function. Every shape deformation that attracts the misclassified pixels towards the true contour seems inevitably to move correctly identified pixels leading to an increase in the objective function. This is reflected in Table 5.3 in which the comparison between the estimates of the distribution parameters based on the true and reconstructed shape are reported. We note that we obtained good estimates of μ_E and σ_E, while the estimate of the other parameters are clearly more affected by the



(a) (b)

Figure 5.17: *(a) Data generated from the true shape shown in Fig. 5.16 by filling its interior and exterior with values drawn from* $\mathrm{N}\left(6,(0.5)^2\right)$ *and* $\mathrm{N}\left(2,(0.5)^2\right)$ *distributions respectively. (b) The reconstructed (green) and true (red) lesion contour.*

Case 3			Estimated value based on the true shape	Estimated value based on reconstruction
SNR=18 dB	Region I	μ_{I}	5.88	5.05
		σ_{I}	0.49	1.65
	Region E	μ_{E}	1.99	2.01
		σ_{E}	0.51	0.60

Table 5.3: *Estimates of the distribution parameters based on the true and reconstructed shape from the data shown in Fig. 5.17(a).*

misclassified pixels.

Case 4

In order to improve the results shown in Fig. 5.17(b), we modified our matching algorithm. The new version, that locks pixels identified as marking the discontinuity as described in Section 5.5, tries to overcome the limitation seen in Case 3. We do not present results for the first type of simulated data because they were almost identical to the ones already shown. In Fig.5.18 we show the reconstruction of the dataset of Fig. 5.17(a) obtained with the new version of the algorithm. In this case, the percentage of correctly identified contour pixels has increased to 93% from 89%. This reconstruction is preferable to the

Figure 5.18: *The reconstructed (green) and true (red) lesion contour obtained by applying the modified algorithm to the data shown in Fig 5.17(a).*

Case 4			Estimated value based on the true shape	Estimated value based on reconstruction
SNR=18 dB	Region I	μ_{I}	5.88	5.74
		σ_{I}	0.49	0.83
	Region E	μ_{E}	1.99	2.03
		σ_{E}	0.51	0.64

Table 5.4: *Estimates of the distribution parameters based on the true and reconstructed shape obtained by the modified algorithm for the data shown in Fig. 5.17(a).*

one shown in Fig. 5.17(b) not only because this percentage is larger, but also because the misclassified pixels are closer to the real contour. In Table 5.4 we report the estimated values of the parameters. Comparing these estimates with those given in Table 5.3, we note that the estimates of μ_{I} and σ_{I} have greatly improved. In addition, the reconstruction was obtained much faster than before. In Fig. 5.19 we compare the values (blue line) of the objective function after each iteration of the original and modified matching algorithms. We also show the value of the scaling parameter at each deformation (red line). From Fig. 5.19(a) we see that 69 iterations of the original algorithm are needed to reach a minimum of the objective function, and this happens when $\alpha^{(32)} = 3.5$. Instead, the new version stops after four iterations and uses just one value of the scaling parameter, namely

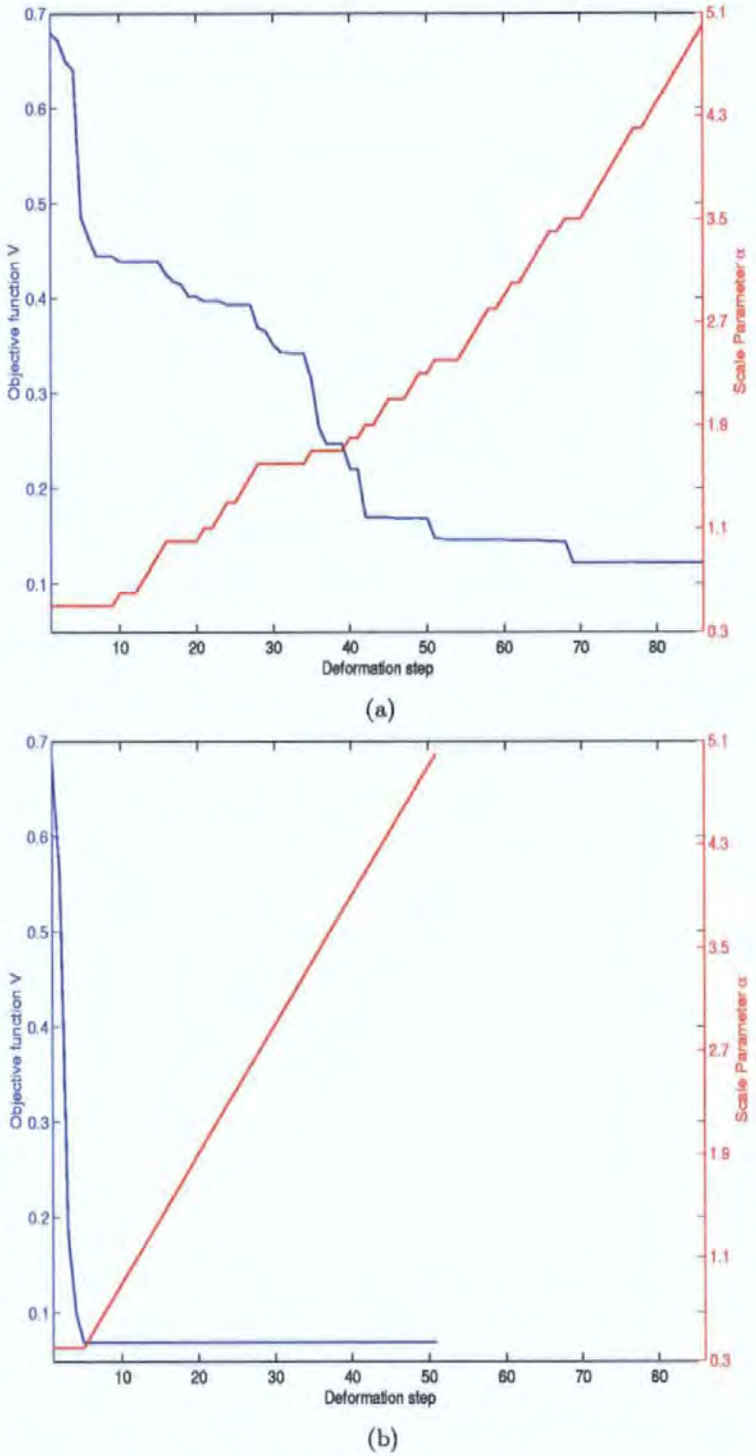


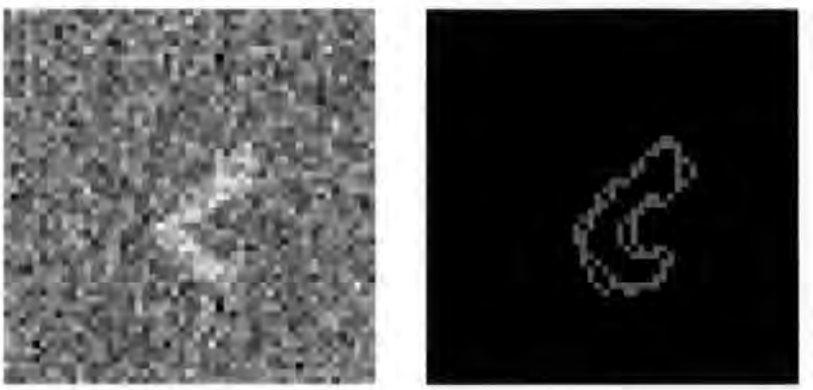
Figure 5.19: *The value of the objective function after each iteration of the matching algorithm (blue curve, left hand axis) and the value of the scaling parameter used for each deformation (red curve, right hand axis) for (a) the original and (b) the modified algorithms. The modified algorithm attains a lower minimum of the objective function after considerably fewer iterations.*

$\alpha^{(1)} = 0.4$. Furthermore, we note that this minimum of the objective function is lower, namely 0.046 compared with 0.138. As an example in Fig. 5.20 we show the four steps of our modified matching algorithm that lead to the reconstruction of Fig. 5.18. In this figure the locked (red) and unlocked (green) pixels at each iteration of the algorithm are reported. We note that in the final optimal shape all the pixels passed the hypothesis test and so are locked.

Figure 5.20: *The locked (red) and unlocked (green) pixels at each iteration of the modified algorithm applied to the data shown in Fig. 5.17(a) and leading to the reconstruction shown in Fig. 5.18.*

Case 5

Since the performance of the algorithm observed so far seems very encouraging, we decided to test our procedure with data characterized by a lower SNR. In Fig. 5.21 we display a dataset with SNR= 6 dB (obtained as described in Case 2), together with the lesion reconstruction obtained using the modified algorithm. In this case 63% of the contour pixels are correctly identified and the reconstructed contour still matches the lesion contour sufficiently well. Moreover, we note from Table 5.5 that good estimates of the parameters have been obtained. As in the previous cases we repeated the analysis with different noise realizations generated in the same way. We observed that the matching algorithm



(a) (b)

Figure 5.21: *(a) Data generated from the true shape shown in Fig. 5.16 by filling its interior and exterior with values drawn from* $\mathrm{N}(3,(0.5)^2)$ *and* $\mathrm{N}(2,(0.5)^2)$ *distributions respectively. (b) The reconstructed (green) and true (red) lesion contour.*

Case 5			Estimated value based on the true shape	Estimated value based on reconstruction
SNR=3 dB	Region I	μ_{I}	2.87	2.74
		σ_{I}	0.47	0.64
	Region E	μ_{E}	2.00	1.98
		σ_{E}	0.49	0.50

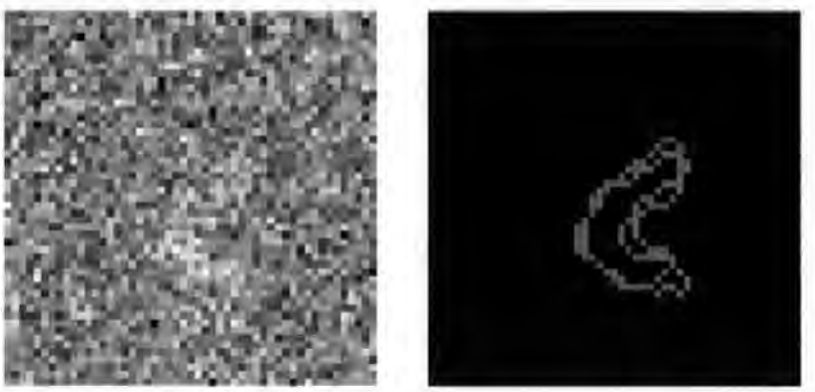
Table 5.5: *Estimates of the distribution parameters based on the true and reconstructed lesion obtained by applying the modified algorithm to the data shown in Fig. 5.21(a).*

produced reliable reconstructions despite the low SNR of the data.

Case 6

Encouraged by the results obtained with very low SNR datasets, we decided to test the algorithm on an extreme case: we adopted the true shape shown in Fig. 5.16 and drew interior and exterior values from $\mathrm{N}(2.5,(0.5)^2)$ and $\mathrm{N}(2,(0.5)^2)$ distributions respectively. This dataset and the corresponding reconstruction are shown in Fig. 5.22. Although the SNR of this dataset is 0 dB, we can see from this figure that the reconstruction is still close to the true lesion. We obtained 53% of correctly identified contour pixels and the estimates of the parameters reported in Table 5.6 are close to the true values. This shows

how powerful the modified algorithm is in reconstructing the lesion contour even with extremely noisy data.



(a) (b)

Figure 5.22: *(a) Data generated from the true shape shown in Fig. 5.16 by filling its interior and exterior with values drawn from* $\mathrm{N}\left(2.5,(0.5)^2\right)$ *and* $\mathrm{N}\left(2,(0.5)^2\right)$ *distributions respectively. (b) The reconstructed (green) and true (red) lesion contour.*

The simulation study has been very important in assessing the potential of the procedure. It allowed us to determine the best performing objective function and to set the parameters such as β and w on which the matching procedure depends. Now that we tuned properly our matching algorithm we can test it on real data.

5.6.2 Real data

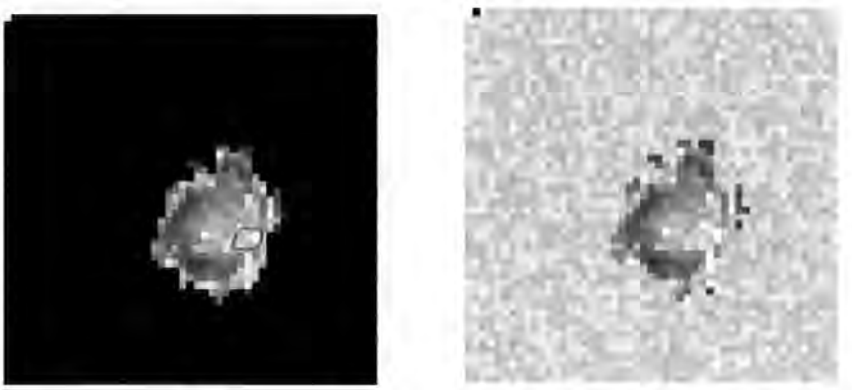
In this section we show the results obtained using real data consisting of the classification attributes described in Section 4. In particular, we use the attributes **A** obtained from

Case 6			Estimated value based on the true shape	Estimated value based on reconstruction
SNR=0 dB	Region I	μ_{I}	2.44	2.42
		σ_{I}	0.49	0.48
	Region E	μ_{E}	1.99	2.01
		σ_{E}	0.48	0.49

Table 5.6: *Estimates of the distribution parameters based on the true and reconstructed lesion obtained by applying the modified algorithm to the data shown in Fig. 5.22(a).*

the non-parametric approach shown in Fig. 4.2 within the tumoral region $\mathcal{I}$. As we described in Section 5.1, the prototype template $\mathbf{T_0}$ is obtained in such a way that it has approximately the same size as $\mathcal{I}$. Now, because of the definition of $\mathrm{E}(\mathbf{T})$, during the matching procedure some pixels of this region may fall outside $\mathcal{I}$. For this reason, we need to consider a region larger than the tumoral region. The dimension of this region has been fixed at 50×50 pixels because this represents a support typically large enough to include all the shapes generated in the matching procedure. The values of the pixels of this support that do not belong to $\mathcal{I}$ have been assigned at random. We now discuss how this is done. Initially, a small region within $\mathcal{I}$ that corresponds to benign tumoral tissue is selected by radiologists. This region is shown in red in Fig. 5.23(a). Then the range of attribute values inside this region is found and pixels outside $\mathcal{I}$ are assigned values drawn from a uniform distribution in this range. A different approach could be to adopt a normal distribution for these attribute values and then estimate its parameters μ_{E} and σ_{E} keeping them fixed in the matching procedure. In this way the lesion reconstruction would depend only on μ_{I} and σ_{I}. We did not choose this approach because we think that the small size of the supervised region within $\mathcal{I}$ could lead to very biased estimates of μ_{E} and σ_{E}. In general, given that $\mathrm{E}(\mathbf{T})$ represents the smallest rectangular region containing $\mathbf{T}$ and the template is typically inside the tumoral region, there will always be a smaller proportion of simulated values compared to real ones in $\mathrm{E}(\mathbf{T})$. Because of this, simulated

values should not substantially affect the lesion reconstruction. In Fig. 5.23(b) we display the full dataset, that is the classification attribute within $\mathcal{I}$ together with the simulated data outside $\mathcal{I}$. The modified version of the algorithm was applied to this dataset and



(a) (b)

Figure 5.23: *(a) The classification attributes* **A** *obtained in Section 4.4 restricted to the tumoral region* $\mathcal{I}$*, together with the reference region corresponding to benign tumoral tissue selected by the radiologist (red). (b) The final dataset. The values of the pixels outside* $\mathcal{I}$ *are uniformly distributed over the same range as the values in in the benign tumoral region*

in Fig.5.24 we show the reconstructed lesion superimposed in yellow on the data. In this case, since we do not know the 'true' lesion contour we cannot estimate the accuracy of this reconstruction. However, we may validate this result using the feedback of expert radiologists who can judge the reliability of the reconstructed lesion. We have shown our result to such radiologists who judged the reconstruction to be reliable.

5.7 Conclusions

In this chapter we presented an approach based on deformable templates to classify the pixels of the tumoral lesion. Our method exploits available 'a priori' knowledge about the spatial structure of the lesion contour. This is represented by means of a prototype shape. First, we derived and discussed a deformation model that we use to transform this

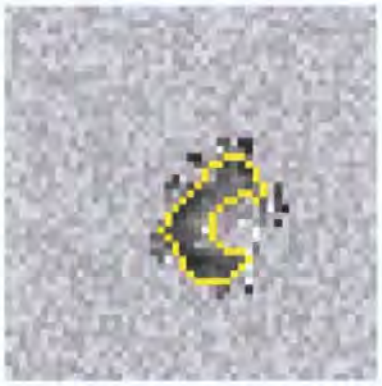


Figure 5.24: *The final lesion reconstruction (yellow) superimposed to the data.*

shape. Then, we studied the richness of the space of shapes generated by this model. The results show that the deformation model is sufficiently versatile to reach shapes typical in this kind of study. Since shapes corresponding to 'extreme' deformations can lose simple connectivity, we developed a new filter to recover this important shape feature. By comparing our results with those obtained using standard morphological filters, we found that our reconstructions generally preserve more information about the spatial structure of the contour.

A new algorithm based on the deformation model and a specially designed objective function was developed to reconstruct the lesion from the classification attributes. In addition, we modified our new filter to identify the interior of a given shape. We tested our algorithm on synthetic and real data. The synthetic data are based on a shape obtained from the classification method of Section 4. In this way this shape reflects a realistic contour for these kinds of tumour. We showed that the proposed algorithm performs well even with extremely noisy datasets. This may be very important since the classification attributes depend on the number of acquired images of the sequence, and so if we decrease this number, these attributes may become more noisy. Reducing the number of acquired images will make the overall acquisition time shorter and this will lead to many advantages

from a clinical point of view such as:

- less time spent by the patient in the tomograph;
- more time available to analyse more patients;
- lower cost for each analysis.

The results obtained from the real dataset cannot be assessed accurately since we do not know the true lesion contour and we do not have any results from histological analysis yet. The radiologists involved in our study judge the classification results reliable. However, we want to stress that this is a preliminary result that must be confirmed by future work. It is our intention, as we explain in detail in Section 7, to extend our methodology to use more than one classification attribute and to adopt a Bayesian approach for the optimization of the objective function. To identify the prototype shape more accurately, we plan to involve doctors in analysing a large number of patients using the software BanDITS described in Section 6. In this way, we will build a library of prototype templates based on classification results from a large sample.

Chapter 6

BAnDITS: a Software Package for the Analysis of dMRI Sequences

At this stage of our investigation feedback from radiologists is crucial to validate the restoration and classification results. In fact, we think that a clear confirmation of the validity of our results could only come from a survey conducted on a large sample of patients. Furthermore, a large scale survey could provide us with a library of suitable prototype templates for the deformable template classification approach presented in Chapter 5. In order to involve more doctors in our research, since they are non-expert in Bayesian analysis we developed a software package that implements the methods described in this thesis. We called this software BAnDITS (Bayesian Analysis of Dynamic Images for Tumour Studies). BAnDITS has a user friendly interface and does not require any knowledge of the underlying methodology. In this chapter we describe briefly the main features of this software. In Fig. 6.1 we show the BAnDITS opening screen. As we can see a series of pop-up menus allows the user to choose between many different operations. These can be grouped into four main categories:

- dMRI sequence import;
- image visualization;

Figure 6.1: *BAnDITS opening screen. A series of pop-up menus allows the user to perform different operations.*

- image restoration;
- image classification.

We now discuss some details of each category.

dMRI sequence import

BAnDITS is able to import dMRI sequences directly from the tomograph. The original DICOM data are converted into a format more appropriate for the analyses, namely a `MATLAB` .mat format. For every new patient the images are stored in a folder, the location and the name of which can be chosen by the user.

Image visualization

Once a patient has been selected, the user can visualize the original and restored dMRI sequence, the original and restored ROI sequence, the current classification attributes and classification image. Moreover, a reference image can be selected (such as the classification image, an image of the restored sequence or the classification attributes) and the temporal evolution of pixels within that image can be analyzed. This helps the user to interpret both the restoration and classification results.

Image restoration

First, a ROI for the chosen patient is selected. We can either select a new ROI and perform the hypothesis test to find the tumoral region $\mathcal{I}$, or choose a previously selected ROI where the complete dMRI analysis has already been performed. In this case the user may want to repeat the image restoration and classification steps using a different approach in order to compare the new results with those previously obtained. This step is shown in Fig.6.2.

The ability to select different ROIs within the same breast allows us to analyse patients

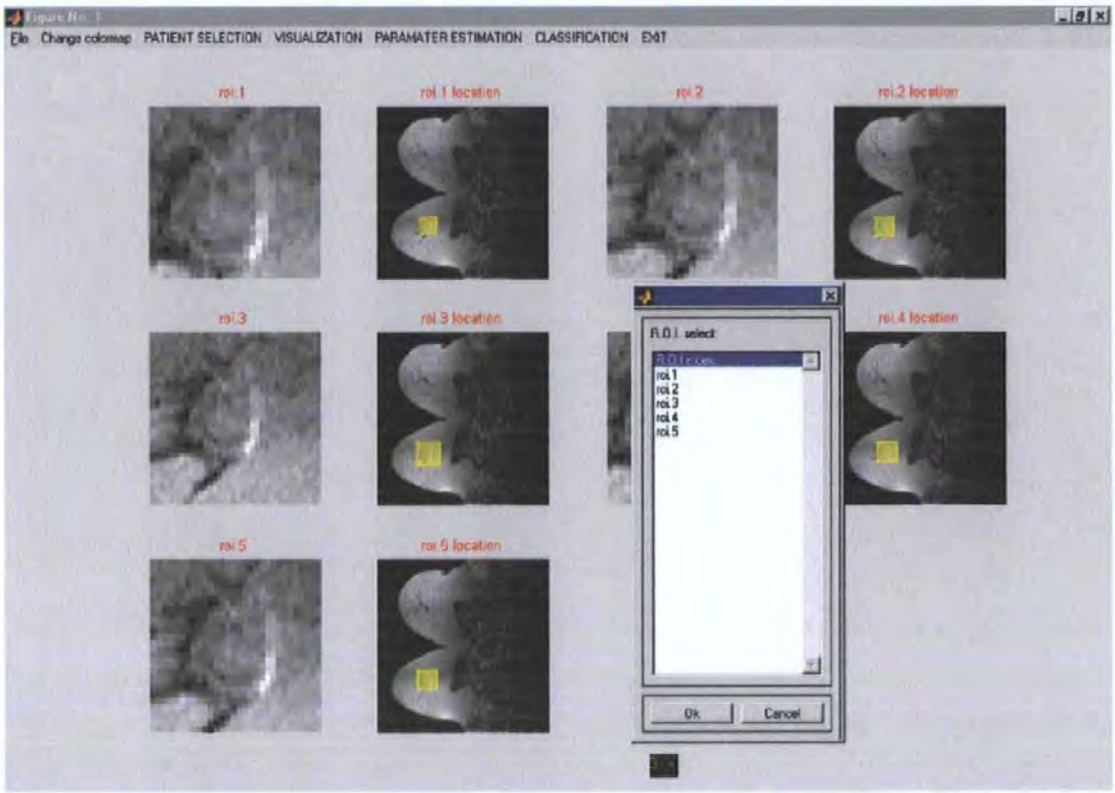


Figure 6.2: *The ROI selection step. For each patient different ROIs can be selected. In this way more than one lesion in the same breast can be analysed.*

with more than one lesion. Once $\mathcal{I}$ is identified the image restoration step, either using the non-parametric or parametric approach, can be performed. This step includes the automatic estimation of the hyper-parameters.

Image classification

The image classification step can be performed using either the parametric or the non-parametric classification attributes. As we show in Fig. 6.3, once the classification image is obtained, it can be validated by checking the temporal evolution of the classified pixels.

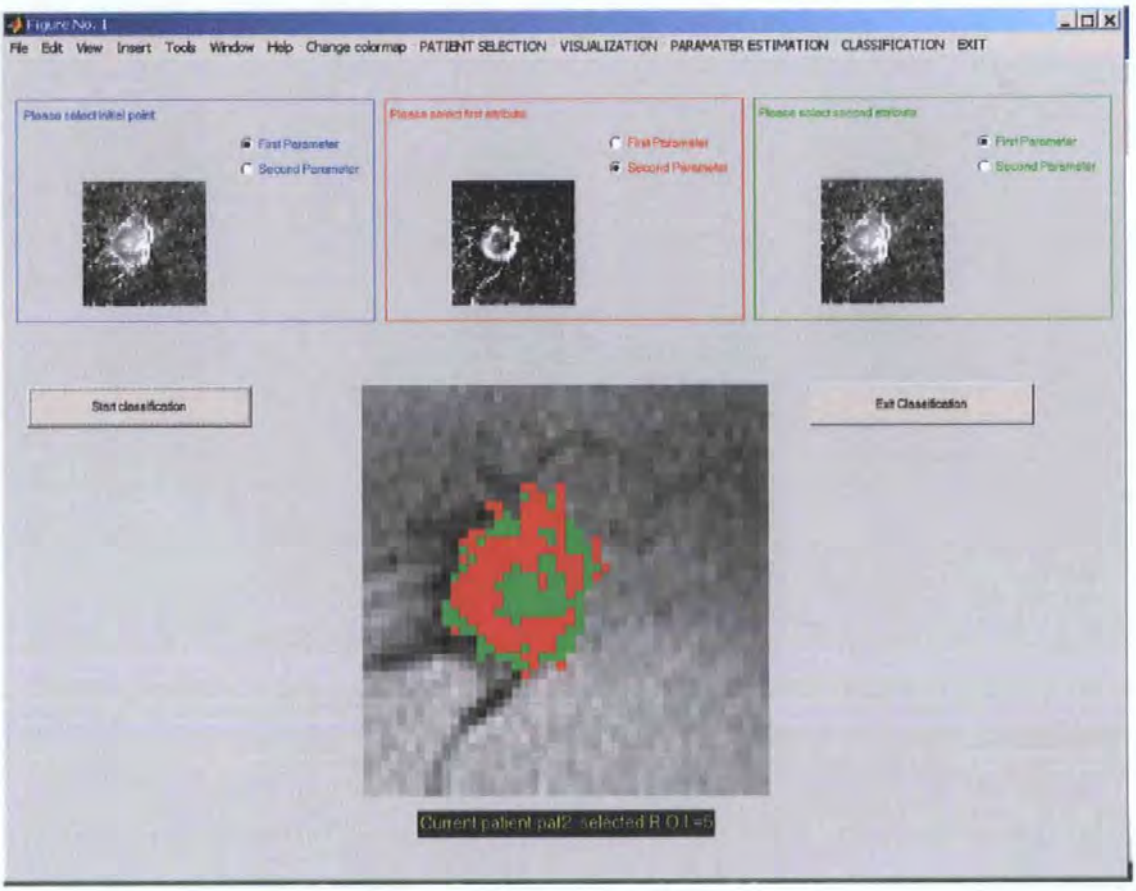


Figure 6.3: *Classification results. The classification image has been obtained selecting the initial point and the two classification attributes.*

BAnDITS has been developed in a **MATLAB** environment by building and linking **FORTRAN** and **C++** subroutines to the main **MATLAB** program. The graphical interface is built in **MATLAB**, while the heavy computations such as the MCMC algorithms are implemented in **FORTRAN** or **C++**. The whole package is compiled as a stand-alone executable. This means that it can be installed on any machine independent of the operating system or whether or not **MATLAB** is installed. In fact, once compiled the stand-alone executable contains all the necessary graphical and computational libraries.

The restoration and classification tasks discussed in this thesis including hyper-parameter estimation take no more than two minutes on a PC with a 2.6 GHz Pentium 4 processor and 256 MB of RAM.

Chapter 7

Discussion and Future Work

In this thesis we present a novel Bayesian methodology and a deformable template based approach for analysing dMRI breast images. The analysis of these images consists of two tasks: image restoration and image classification. Since one of the disadvantages of Bayesian methods is their heavy computational burden we address this issue by proposing a method to restrict our analyses to a smaller region of the breast (called the tumoral region) and by developing a specially tailored Metropolis-Hastings algorithm. The identification of the tumoral region considerably reduces the number of pixels to be analyzed, while our modified version of the Metropolis-Hastings algorithm with its special proposal distribution shows higher speed of convergence than algorithms of the same family with more commonly used proposals. In this way, the overall time for analysing a typical dMRI sequence is reduced to no more than two minutes using a PC with a 2.6 GHz Pentium 4 processor and 256 MB of RAM.

The problem of estimating hyper-parameters of the prior models has been addressed by presenting criteria to obtain them. Hence, our Bayesian methodology becomes automatic, apart from the selection of the ROI and the reference region for the hypothesis test by the user.

For the image restoration task we developed a non-parametric and a parametric

approach and we compared the results obtained with the two methods. The image restorations yielded by both methods show that the effect of random distortions are successfully reduced while important information such as the structure of the lesion is preserved. The images of the restored sequences obtained with the two methods seem very similar. However, some differences in the temporal patterns are evident. In particular, the parametric approach sometimes seems to be too rigid and so unable to model the temporal evolution properly. On the other hand the non-parametric approach can produce biased reconstructions. From our analysis the parametric approach seems to perform better within the tumoral region, although this result must be confirmed by a study on a larger number of patients.

The classification results obtained by adopting non-parametric and parametric attributes are very similar. These classification images have been judged reliable by expert radiologists. However, we can only compare the two methods precisely by conducting a broader study on a larger sample of patients. All our results indicate that our new approach is more informative than the standard method previously adopted by physicians. In fact, with our methodology not only is the location of the lesion correctly identified, but also its internal structure becomes evident. This internal structure is not visible from the results obtained from the standard method and knowledge of it can yield important medical information. Most of the above Bayesian methodology is summarized in de Pasquale *et al.* (2003).

In the first part of the work the adopted prior models take into account the temporal continuity and spatial homogeneity between neighbouring pixels. Since the spatial structure of the lesion contour plays an important role in the analysis of dMRI sequences, in the second part of this thesis we developed a method in which higher level prior information about the actual shape of the lesion is taken into account. This method is based on

parametric prototype deformable templates. In this approach an assumed prototype shape is deformed by parametric non-affine transformations. An objective function that depends on the parameters of the transformation is minimized to reconstruct the lesion contour. We defined the appropriate non-affine transformations and developed a multiresolution deformation model based on these. Then we investigated the richness of the space of shapes generated by this deformation model. This model seemed to be sufficiently versatile to reach shapes that are characteristic for these studies. We implemented a matching algorithm based on the deformation model and developed a novel objective function. We tested our procedure on synthetic data with a variety of SNRs. The results show that our matching algorithm performs well in reconstructing the lesion contour even with extreme low SNRs. The simulation study was very important as it helped us to assess the objective function and the performance of the algorithm. Using the version of the algorithm that performed best with simulated data, we analyzed a real data set. In this case we cannot of course judge the lesion reconstruction as we did for simulated data since we do not know the true lesion contour. However, the shape reconstruction that we obtained was judged to be reasonable by radiologists.

As part of the development of our deformable template based approach we proposed a new filter that restores the connectivity without smoothing the shape. In addition, we modified our filter to identify the interior of a given shape.

There are many aspects of this work which we plan to develop further.

For the Bayesian methodology, discussed in Chapters 3 and 4, future work would involve the generalization of the methodology to deal with three-dimensional dMRI sequences. In addition, the integration of the models and methods used in the restoration and classification steps would be studied. It would also be important to validate the restoration and classification results by means of a study on a large sample of patients.

We aim to do this by giving our software package BAnDITS to a team of physicians who could test the methodology that we have developed on such a large sample. This would further validate our procedure and show its limitations, leading us to prefer one approach over another.

For the template based matching procedure discussed in Chapter 5, further work would involve extending the algorithm to use more than one classification attribute to reconstruct the lesion contour. We would also plan to validate the algorithm using the study on a large sample of patients, as mentioned above. This would allow us also to build a library of prototype templates. From this survey a database of template prototypes for the lesion shapes could also become available for the template based matching procedure. Furthermore, the procedure could be cast into a Bayesian framework by assigning probability models to the deformation parameters and performing the objective function optimization by means of MCMC methods such as simulated annealing.

Appendix A

Estimating the Best Fitting Ellipse of a Shape from its Moments

In Section 5.1 we describe how the prototype template is initially scaled to have roughly the same size as the lesion. This step is based on the estimates of the best fitting ellipses for the tumoral region contour and the prototype template. We describe here how these ellipses are obtained using a procedure based on the definition of the moments of a shape (Hu, 1962). The basic assumption is that the best fitting ellipse and the considered shape share the same moments. Hence, we begin by estimating the moments of the shape. Then, based on these estimates the parameters of the ellipse can be calculated. First of all, let us introduce an ellipse with centroid (x_0, y_0) at the origin. Let us begin by assuming that the angle θ between the major axis of the ellipse and the x axis is zero. In this way the ellipse can be thought of as not rotated. This ellipse can be either represented as: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, where a and b are the length of the major and minor semiaxis respectively or, using a parametric representation, as:

$$\begin{aligned} x &= a\cos t \\ y &= b\sin t, \end{aligned} \tag{A.1}$$

where $t \in [0, 2\pi)$.

We now consider the problem of estimating the parameters of a generic rotated ellipse (a, b, x_0, y_0, θ) from its moments.

First of all, we need to introduce the moments of a generic shape S. The $(p, q)^{\text{th}}$ order moment of S is defined as:

$$\mu_{pq} = \int_S x^p y^q dx dy. \tag{A.2}$$

The collection of moments μ_{pq} can provide useful features of S. In particular:

- μ_{00} measures the area of the shape;
- μ_{20} measures the spread of the shape in the x direction;
- μ_{02} measures the spread of the shape in the y direction.

Moreover, the centroid of S can be found from these moments as $(x_0, y_0) = \left(\frac{\mu_{10}}{\mu_{00}}, \frac{\mu_{01}}{\mu_{00}}\right)$. Using this we can translate S so that its new centroid is at the origin.

Since the moments of a shape contain important information about some of its characteristic features, they may be excellent parameters to use to estimate the best fitting ellipse. Assuming that this ellipse has the same moments as S means that these shapes will share some important features such as area and spread along the axes. From this assumption it is easy to estimate the parameters of the ellipse. In fact, if we substitute the parametric representation of the ellipse (A.1) in (A.2) we have:

$$\mu_{20} = \int_{\text{ellipse}} x^2 dx dy = 2a^3 b \int_0^{\pi} \cos^2 t \sin^2 t dt = \frac{\pi}{4} a^3 b$$

$$\mu_{02} = \int_{\text{ellipse}} y^2 dx dy = 2ab^3 \int_0^{\pi} \cos^2 t \sin^2 t dt = \frac{\pi}{4} b^3 a.$$

We now obtain the length of the major and minor semiaxes a and b as:

$$\begin{aligned} a &= \left(\frac{4}{\pi}\right)^{\frac{1}{4}} \left(\frac{\mu_{20}^3}{\mu_{02}}\right)^{\frac{1}{8}} \\ b &= \left(\frac{4}{\pi}\right)^{\frac{1}{4}} \left(\frac{\mu_{02}^3}{\mu_{20}}\right)^{\frac{1}{8}}. \end{aligned} \tag{A.3}$$

This analysis can be extended to the case where the ellipse is rotated by an angle θ. In this case expressions for the rotated first and second order moments in terms of θ and the previous first and second order moments can be found. The resulting equations can be solved for a, b and θ.

In our study we do not estimate θ because rotations are not considered in the first affine transformation based stage of our matching algorithm. For the lesion contour and the prototype template we first calculate the second order moments using (A.2) and then we estimate the major and minor semiaxis of the best fitting ellipses using (A.3). The ratios between the corresponding axes of these best fitting ellipses are calculated, so yielding scaling factors for the prototype template. Finally, the centroids of the two shapes are made to coincide.

Appendix B

Morphological Operations

B.1 Skeletonisation

In many fields of image analysis such as pattern recognition and shape analysis, methods for extracting some object features such as contours are needed. Among many different approaches one of the most popular is to use mathematical morphology to transform the object into a set of idealised thin lines which condense the information about the object while preserving its homotopy (for a definition of homotopy see Soille, 1999). The corresponding morphological transformations are called Skeletonisations or medial axis transformations. Several formal definitions of the skeleton are available for continuous space. Some of these are based on grass-fire or wavefront propagations, distance functions, maximal disks or minimal paths, and lead to the same thin lines. The extension of these skeleton definitions from continuous to discrete space is not straightforward, with different methods leading to different discrete skeletons. In this appendix we describe a particular Skeletonisation for a binary image based on homotopic thinnings. We refer the reader to Soille (1998) for a complete review.

Homotopic Thinnings

The Skeletonisation that we adopt is based on morphological transformations called Homotopic Thinnings. We chose this set of transformations because the resulting skeletons preserve important features of the original shape such as simple connectivity. In order to explain the details of our Skeletonisation we need to introduce the Hit or Miss transform and the Thinning transform.

The basic idea behind the Hit or Miss transform consists of extracting image pixels of a binary image that have a particular neighbouring configuration. The neighbouring configuration is described in terms of two disjoint sets. The first set represents the configuration of neighbouring pixels that have the value one in the binary image. The second set represents the configuration of zero valued pixels. These two sets form a composite structuring element $B = (B_1, B_2)$. The origin of B is the central pixel. An example of a composite structuring element is shown in Fig. B.1. Each pixel of the image is visited in turn and the following question is posed: 'Does the composite structuring element with origin at this pixel fit the image?'. If the answer is affirmative, the visited pixel is given the value one. In the following, a Hit or Miss transformation of a set X by a composite structuring element B will be indicated by $HMT_B(X)$.

Now that we introduced the Hit or Miss transformation we are able to define the Thinning transform. A particular Thinning transform in which homotopic[1] composite structuring elements are used is called Homotopic Thinning.

The Thinning transform consists of removing shape pixels having a particular configuration. This is achieved by subtracting the Hit or Miss transform from the original image. Hence, the Thinning of a binary image is denoted by $X \circ B$ and defined as the set

[1] A composite structuring element is said to be homotopic if the number of its black and white connected components is not modified when modifying the state of its origin.

difference:

$$X \circ B = X \setminus HMT_B(X). \tag{B.1}$$

Finally, the skeleton is obtained by Thinning the image by a series of homotopic composite structuring elements and their rotations until stability is reached, i.e. the input set has been reduced to a set of thinned lines. The Thinnings are performed according to a sequence based on all rotations of the homotopic composite structuring elements. This process is called sequential thinning and is denoted by $\underline{\circ}$. We have therefore the following formula for the sequential thinning of a set X by the n rotations $\theta_1 B, \theta_2 B, \ldots, \theta_n B$ of the homotopic structuring elements:

$$X \underline{\circ} B = (\ldots((X \circ \theta_1 B) \circ \theta_2 B) \circ \ldots) \circ \theta_n B.$$

In fact, the set $X \underline{\circ} B$ is the skeleton of X under the chosen composite structuring element B. A different skeleton is obtained for different choices of the homotopic composite structuring element. In particular, in this study to obtain eight-connected skeletons we adopted the two composite structuring elements shown in Fig. B.1 and their four rotations. In this

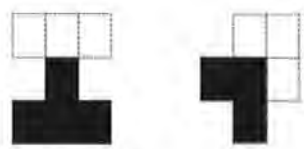


Figure B.1: *The composite structuring elements B adopted to obtain eight connected skeletons. The origin is the central pixel. The white and black pixels represent the sets B_1 and B_2 respectively.*

figure B_1 and B_2 are the sets of white and black pixels respectively.

Because the Skeletonisation is repeated as many times as the shape deformations in our matching algorithm, it can be thought of as part of the deformation itself and so it is vital to implement this transformation in a very fast way in order not to slow down the procedure. To apply the sequence of homotopic thinnings to every shape pixel of the

image is very expensive from a computational point of view. A more efficient way to implement this transformation is to generate a Look Up Table that contains the result of the Skeletonisation for every possible neighbouring configuration. In other words, we give a stack number to every possible neighbouring configuration which becomes the input address of the Look Up Table. Hence the adopted Skeletonisation proceeds by visiting each shape pixel. The neighbouring configuration and its corresponding stack number are identified and the value of the skeletonized shape pixel is then immediately given by the Look Up Table. To associate a stack number to a neighbouring configuration is fairly simple. In fact, it is sufficient to adopt a fixed visiting order to the neighbouring pixels, so that the configuration becomes a pattern, and then assign a stack number to every possible pattern. In Fig. B.2 we show the visiting order of the neighbouring pixels. For

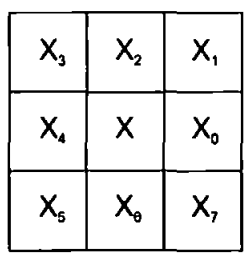
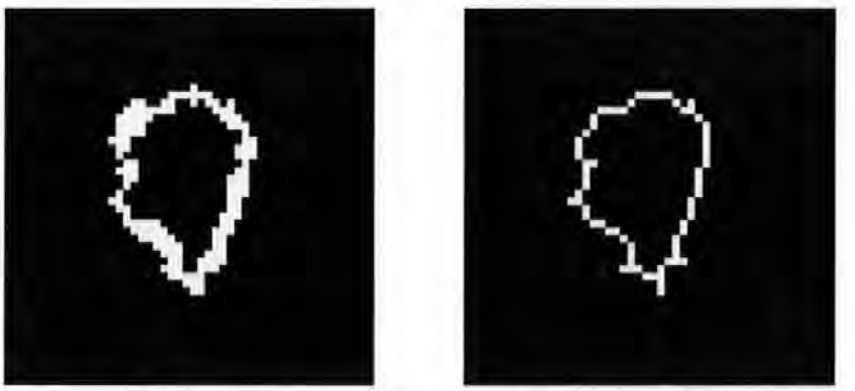


Figure B.2: *The visiting order for neighbouring pixels. For every shape pixel X the corresponding pattern is* $(X_0, X_1, X_2, X_3, X_4, X_5, X_6, X_7)$. *We note that we have* $2^8 = 256$ *possible patterns.*

every shape pixel X the corresponding pattern is $(X_0, X_1, X_2, X_3, X_4, X_5, X_6, X_7)$. We note that we have $2^8 = 256$ possible patterns. As example in Fig B.3 we show the effect of Skeletonisation on a particular shape. As we can see from this figure the original shape has been reduced to a set of idealised thin lines. We note that the skeleton is simply connected.



Original shape Skeletonized shape

Figure B.3: *The effect of Skeletonisation. The original shape has been reduced to idealised thin lines preserving simple connectivity.*

B.2 Bridge transformation

As we mentioned in Section 5.2.1, some extreme deformations of our model do not preserve the connectivity of the shape. In our approach this characteristic of the shape is fundamental. Hence, in order to recover shape connectivity we employed the morphological Bridge transformation. This transformation is able to connect neighbouring shape pixels previously unconnected. Following the notation introduced in Section B.1, we have that the Bridge transformation changes X into X' according to:

$$X' = X \cup (P_1 \cup P_2 \cup \cdots \cup P_6),$$

where

$$\begin{aligned}
P_1 &= \overline{X}_2 \cap \overline{X}_6 \cap (X_3 \cup X_4 \cup X_5) \cap (X_0 \cup X_1 \cup X_7) \cap \overline{P}_Q \\
P_2 &= \overline{X}_0 \cap \overline{X}_4 \cap (X_1 \cup X_2 \cup X_3) \cap (X_5 \cup X_6 \cup X_7) \cap \overline{P}_Q \\
P_3 &= \overline{X}_0 \cap \overline{X}_6 \cap X_7 \cap (X_2 \cup X_3 \cup X_4) \\
P_4 &= \overline{X}_0 \cap \overline{X}_2 \cap X_1 \cap (X_4 \cup X_5 \cup X_6) \\
P_5 &= \overline{X}_2 \cap \overline{X}_4 \cap X_3 \cap (X_0 \cup X_6 \cup X_7)
\end{aligned}$$

$$P_6 = \overline{X}_4 \cap \overline{X}_6 \cap X_5 \cap (X_0 \cup X_1 \cup X_2)$$

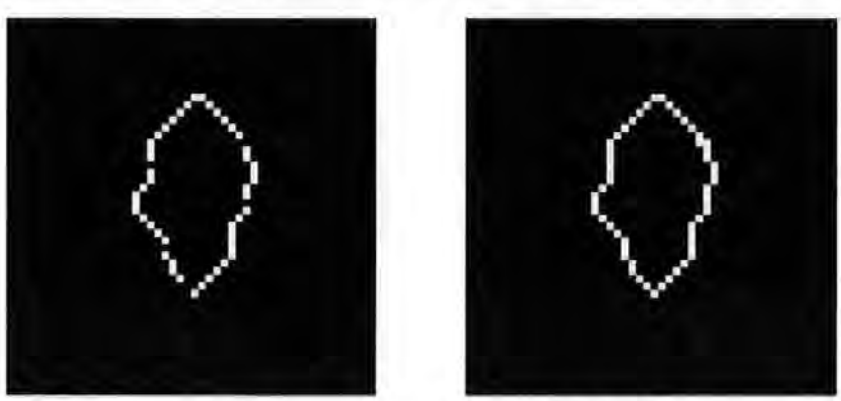
in which

$$P_Q = L_1 \cup L_2 \cup L_3 \cup L_4 \tag{B.2}$$

where

$$\begin{aligned} L_1 &= \overline{X} \cap \overline{X}_0 \cap X_1 \cap \overline{X}_2 \cap X_3 \cap \overline{X}_4 \cap \overline{X}_5 \cap \overline{X}_6 \cap \overline{X}_7 \\ L_2 &= \overline{X} \cap \overline{X}_0 \cap \overline{X}_1 \cap \overline{X}_2 \cap X_3 \cap \overline{X}_4 \cap X_5 \cap \overline{X}_6 \cap \overline{X}_7 \\ L_3 &= \overline{X} \cap \overline{X}_0 \cap \overline{X}_1 \cap \overline{X}_2 \cap \overline{X}_3 \cap \overline{X}_4 \cap X_5 \cap \overline{X}_6 \cap X_7 \\ L_4 &= \overline{X} \cap \overline{X}_0 \cap X_1 \cap \overline{X}_2 \cap \overline{X}_3 \cap \overline{X}_4 \cap \overline{X}_5 \cap \overline{X}_6 \cap X_7. \end{aligned}$$

There are 119 cases in which the Bridge transformation changes 0 into 1 so connecting previously unconnected pixels. As an example we show in Fig. B.4 the effect of the Bridge transformation. As we can see, in the final image previously unconnected shape pixel have been connected by this transformation. The drawback of this morphological operation is



Original shape Bridged shape

Figure B.4: *The effect of the Bridge transformation. Previously unconnected shape pixels have been connected.*

that it tends to make narrow shapes multiple connected. In fact, when the edges of the

considered shape are separated by just one pixel, the Bridge operation will connect them.

Bibliography

[1] Abramovich, F., Bailey, T.C. and Sapatinas T. (2000) Wavelet analysis and its statistical applications. *J. Roy. Statist. Soc.* Ser. D, **49**, 1–29.

[2] Amit, A., Grenander U. and Piccioni M. (1991) Structural image reconstruction through deformable templates. *J. Amer. Statist. Assoc.*, **86**, 376–388.

[3] Ballard, D.H., (1981) Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recogn.*, **13**, 111–122.

[4] Baumberg, A. and Hogg, D. (1995) An adaptive eigenshape model. *Proceedings of the 6th British Machine Vision Conference*, **15**, 617–634.

[5] Besag, J. (1986) On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc.* Ser. B, **48**, 259–302.

[6] Besag, J. (1989) Towards Bayesian image analysis. *J. Appl. Statist.*, **16**, 3, 395–407.

[7] Brix, G., Semmler, W., Schad, L.R., Layer, G. and Lorenz, W. (1991) Pharmacokinetics parameters in CNS Gd-DTPA enhanced MR imaging. *J. Comput. Assist. Tomogr.*, **15**, 621–628.

[8] Butkov, E. (1968) *Mathematical Physics*. MA: Addison-Wesley.

[9] Chin, R.T. and Dyer, C.R. (1986) Model-based recognition in robot vision. *ACM Comput. Surveys*, **18**, 67–108.

[10] Cohen, L.D. and Cohen, I. (1993) Finite-element methods for active contour and balloons for 2D and 3D images. *IEEE Trans. Patt. Anal. Mach. Intell.*, **15**, 131–147.

[11] de Pasquale, F., Barone, P., Sebastiani, G. and Stander J. (2003) Bayesian analysis of dynamic Magnetic Resonance breast images. Accepted for publication in *J. Roy. Statist. Soc.* Ser. C.

[12] de Pasquale F., Sebastiani G., Egger E., Guidoni L., Luciani A.M., Marzola P., Manfredi R., Pacilio M., Piermattei A., Viti V. and Barone P. (2000) Bayesian estimation of relaxation times T1 in MR images of irradiated Fricke-agarose gels. *Magn. Reson. Imaging*, **18**, 721–731.

[13] Donoho and Johnstone, I. M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

[14] Donoho and Johnstone, I. M. (1995) Adapting to unkown smoothness via wavelet wavelet shrinkage. *J. Am. Stat. Ass.*, **90**, 1200–1224.

[15] Dryden, I.L. and Mardia, K.V. (1998) *Statistical Shape Analysis*. New York: John Wiley and Sons.

[16] Ferlay, J., Bray F., Pisani, P. and Parkin, D.M. (2001) GLOBOCAN 2000: Cancer incidence, mortality and prevalence worldwide. *IARC CancerBase*, **5**.

[17] Gelman, A. and Meng, X.L. (1998) Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.*, **13**, 163–185.

[18] Geman, S., and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.*, **6**, 721–741.

[19] Geman, S. and McClure, D. (1987) Statistical methods for tomographic image reconstruction. *Int. Stat. Inst. Bullettin*, **52**, 4–20.

[20] Glad, I.K. and Sebastiani, G. (1995) A Bayesian approach to syntethic magnetic resonance imaging. *Biometrika*, **82**, 237–250.

[21] Glasbey, C.A. and Horgan, G.W. (1995) *Image Analysis for the Biological Sciences.* New York: John Wiley and Sons.

[22] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall.

[23] Green, P.J. and Richardson, S. (2002) Hidden Markov models and disease mapping. *J. Amer. Statist. Assoc.*, **97**, 1055–1070.

[24] Grenander, U., Keenan, D.M., Mardia, K.V. and Kanji, G.K. (1993) Towards automated image understanding. *Advances in Applied Statistics: Statistics and Images (1)*, **6**, 67–108. Carfax Publishing Company, Abingdon, UK.

[25] Gribbestad, I., Nilsen, G. and Fjosne, H. (1992) Contrast-enhanced magnetic resonance imaging of the breast. *Acta Oncol.*, **31**, 833–842.

[26] Haddadnia, J., Faez, K. and Moallem, P. (2001) Neural network based face recognition with moment invariants. *International Conference on Image Processing, Vol I*, 1018–1021.

[27] Hayton, P.M., Brady M., Smith S. and Moore, N. (1999) A non-rigid registration algorithm for dynamic breast MR images. *Artif. Intell.*, **114**, 125–156.

[28] Hayton, P.M., Brady, M., Tarassenko, L. and Moore, N. (1996) Analysis of dynamic MR breast images using a model of contrast enhancement. *Med. Image Anal.*, **1**, 207–224.

[29] Henkelman, R.M. (1985) Measurement of signal intensities in the presence of noise in MR images. *Med. Phys.*, **12**, 232–233.

[30] Heywang-Kobrunner, S.H. and Beck, R. (1995) *Contrast Enhanced MRI of the Breast.* Second edition. Berlin Heidelberg: Springer Verlag.

[31] Heywang, S.H., Hahn, D., Schmidt, H., Krischke, I., Eiermann, D., Basserman, R. and Lissner, J. (1986) MR imaging of the breast using Gadolinium-DTPA. *J. Comput. Assist. Tomogr.*, **10**, 199–204.

[32] Highnam, R. and Brady, J. (1999) *Mammographic Image Analysis.* Dorderecht: Kluwer Academic Publishers.

[33] Hough, P.V.C. (1962) A method and means for recognizing complex patterns. *US patent number 3,069,654.*

[34] Hu, M.K. (1962) Visual pattern recognition by moment invariants. *IEEE Trans. Inform. Theory,* **8**, 179–187.

[35] Hurn, M.A. (1998) Confocal fluorescence microscopy of leaf cells: an application of Bayesian image analysis. *J. Roy. Statist. Soc.* Ser C, **47**, 361–377.

[36] Hurn, M.A. and Jennison, C. (1995) A study of simulating annealing and a revised cascade algorithm for image reconstruction. *Statist. Comput.*, **5**, 175–190.

[37] Illingworth, J. and Kitter, J. (1988) A survey of Hough Transform. *CVGIP*, **44**, 87–116.

[38] Jain, A.K., Zhong, Y. and Dubuisson-Jolly, M.P. (1988) Deformable template models: A review. *Signal Processing*, **71**, 109–129.

[39] Jain, A.K., Zhong, Y. and Lakshmanan S. (1996) Object matching using deformable templates. *IEEE Trans. Patt. Anal. Mach. Intell.*, **18**, 267–277.

[40] Kass, M., Witkin, A. and Terzopoulos, D. (1988) Snakes: active contour models. *Int. Journal of Computer Vision*, **1**, 321–331.

[41] Krishnan, S., Chenevert, T., Helvie, M. and Londy, F. (1999) Linear motion correction in three dimensions applied to dynamic gadolinium enhanced breast imaging. *Med. Phys.*, **26**, 707–714.

[42] Krzanowski, W.J. (1990) *Principles of Multivariate Analysis.* Oxford: Oxford Science Publications.

[43] Kuhl, C.K., Mielcareck, P., Klaschik, S., Leutner, C., Wardelmann, E., Gieseke, J. and Schild, H. (1999) Dynamic breast MR imaging: are signal intensity time course data useful for differetiating diagnosis of enhancing lesions? *Radiology*, **211**, 101–110.

[44] Leavers, V.F. (1993) Which Hough transform? *CVGIP: Image Understanding*, **58**, 250–264.

[45] Lemarie, F. and Levine, M. (1993) Tracking deformable objects in the plane using active contour model. *IEEE Trans. Patt. Anal. and Mach. Intell.*, **15**, 617–634.

[46] Mallat, S.G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intell.*, **11**, 674–693.

[47] Matthaei, D., Haase, A., Henrich, D. and Duhmke E. (1992) Fast inversion recovery T1 contrast and chemical-shift contrast in high-resolution snapshot flash MR images. *Magn. Reson. Imaging*, **10**, 1–6.

[48] Metropolis, N., Rosenbluth, A.W., Rosenbluth. M.N., Teller A.H. and Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–91.

[49] Mussurakis, S., Buckley, D.L. and Horsman, A. (1997) Dynamic MRI of invasive breast cancer: assessment of three region of interest analysis methods. *J. Comput. Assist. Tomogr.*, **21**, 431–438.

[50] Mussurakis, S., Gibbs, P. and Horsman, A. (1998) Peripherical enhancement and spatial contrast uptake heterogeneity of primary breast tumours: quantitative assesment with dynamic MRI. *J. Comput. Assist. Tomogr.*, **22**, 35–45.

[51] Potts, R. (1952) Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*, **48**, 106–109.

[52] Sebastiani, G. and Godtliebsen, F. (1997) On the use of Gibbs priors for Bayesian image restoration. *Sign. Process.*, **56**, 111–118.

[53] Sebastiani, G. and Sørbye, S. (2002) A Bayesian method for multispectral image data classification. *J. Nonparametr. Statist.*, **14**, 169–180.

[54] Soille, P. (1999) *Morphological Image Analysis.* Berlin: Springer-Verlag.

[55] Sijbers, J., den Dekker, A.J., Raman, E. and Van Dyck, D. (1999) Parameter estimation from magnitude MR images. *Int. J. Imag. Syst. Tech.*, **10**, 109–114.

[56] Stander, J. and Silverman, B.W. (1994) Temperature schedules for simulated annealing. *Statist. Comput.*, **4**, 21–32.

[57] Stuart, M. and Young, M.D. (1988) *Magnetic Resonance Imaging: Basic Principles.* Springer-Verlag: New York.

[58] Szabó, B.K., Aspelin, P., Wiberg, M.K. and Bone, B. (2003) Dynamic MR imaging of the breast - Analysis of kinetic and morphological diagnostic criteria. *Acta Radiol.*, **44**(4), 379–386.

[59] Terzopolous, D., Witkin, A. and Kass, M. (1988) Contraints on deformable models: Recovering 3D shape and non-rigid motion. *Artif. Intell.*, **36**, 91–123.

[60] Tofts, P. and Kermode, A. (1991) Measuremnt of blood brain barrier and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magnet. Reson. Med.* **17**, 357–367.

[61] Weinstein, D., Strano, S., Cohen, P., Fields, S., Gomori, J.M. and Degani, H. (1999) Mapping pathophysiologic features of breast firboadenoma by the three time point (3 TP) contrast enhanced MRI method. *Radiology,* **210**, 233–240.

[62] Winkler, G. (1995) *Image Analysis, Random Fields and Dynamic Monte Carlo Methods.* New York: Springer Verlag.

[63] Yuille, A.L., Hallinan, P.W. and Cohen, D.S. (1992) Feature extraction from faces using deformable templates. *Int. J. Comput. Vision*, **8**, 133–144.