UNIVERSITAT DE GIRONA
ESCOLA POLITÈCNICA SUPERIOR
Departament d'Arquitectura i Tecnologia de Computadors (ATC)

Màster Universitari en Informàtica Industrial, Automàtica,
Computació i Sistemes
(MIIACS)

Atlas-based segmentation of brain MRI
Application to multiple sclerosis

Submitted by
Mariano Cabezas Grebol

Supervisors:
Dr. Xavier Lladó
Dr. Arnau Oliver
Universitat de Girona

Agraïments

Un any més, puc mirar endarrere i veure tot el treball realitzat amb aquestes imatges que fa uns anys m'eren tan desconegudes. Un any més dono les gràcies al grup VICOROB per donar-me suport i ajudar-me a tirar endavant tota una tesi de màster! Durant aquests últims anys, gràcies als meus directors, l'Arnau i en Xavier, a experts externs, la Meritxell, i als diferents doctors i radiòlegs dels hospitals, en Kai, en Lluís i l'Àlex, he après una mica més sobre el nostre estimat cervell i la malaltia de l'esclerosi múltiple. He conegut una mica més d'aprop el problema de la segmentació de ressonàncies magnètiques del "cap", i fins i tot, assistir a seminaris especialitzats en el tema i reunions, sempre acompanyat del cap de grup, en Jordi.

Però això no es tot. Un any més he pogut gaudir de la bona companyia al laboratori amb els meus companys, i sobretot, al meu company dintre dels projectes SALEM, l'Onur i els *cafeters* que com jo han hagut d'estar patint la redacció de la tesi, el projecte o les assignatures pendents: en Pablo, en Pla, en Massi, en Carles, l'Albert i en Gerard (tot i que aquests dos útims passessin tot un semestre fora). Durant aquest últim any, he pogut gaudir també de la companyia de la meva xicota, la Tina, amb la qual he passat molt bons moments i he pogut desconectar (requisit indispensable en tot gran projecte).

A tots vosaltres us dono les gràcies per haver-me aportat alguna cosa. Sense vosaltres aquesta memòria no existiria.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Multiple sclerosis (MS) is the most frequent non-traumatic neurological disease that causes more disability in young adults. It is relatively common in Europe, the United States, Canada, New Zealand, and parts of Australia, but rare in Asia, and in the tropics and subtropics of all continents. Within regions having a temperate climate, the incidence and prevalence of MS increase with latitude, both north and south of the equator. Multiple sclerosis is between twice and three-times more common in women than in men; men have a tendency for later disease onset with a poorer prognosis. The incidence of MS is low in childhood, increases rapidly after the age of 18, reaches a peak between 25 and 35, and then slowly declines, becoming rare at 50 and older. The world estimate is 1.3 to 2.5 million cases of MS, with Western Europe having $350,000$ [31]. Prevalence and incidence of MS increases around the world according to the latest epidemiological studies.

This disease is a chronic, persistent inflammatory-demyelinating and degenerative disease of the central nervous system (CNS), characterised pathologically by areas of inflammation, demyelination, axonal loss, and gliosis scattered throughout the CNS, often causing motor, sensorial, vision, coordination, deambulation, and cognitive impairment [24]. Relapses and progression are the two major clinical phenomena of prototypic MS. Relapses are considered the clinical expression of acute focal or multifocal inflammatory demyelination disseminated in the CNS. Remission of symptoms early in the disease is likely the result of remyelination, resolution of inflammation, and compensatory mechanisms such as redistribution of axolemmal sodium channels and cortical plasticity. These recovery mechanisms are less effective after recurrent attacks.

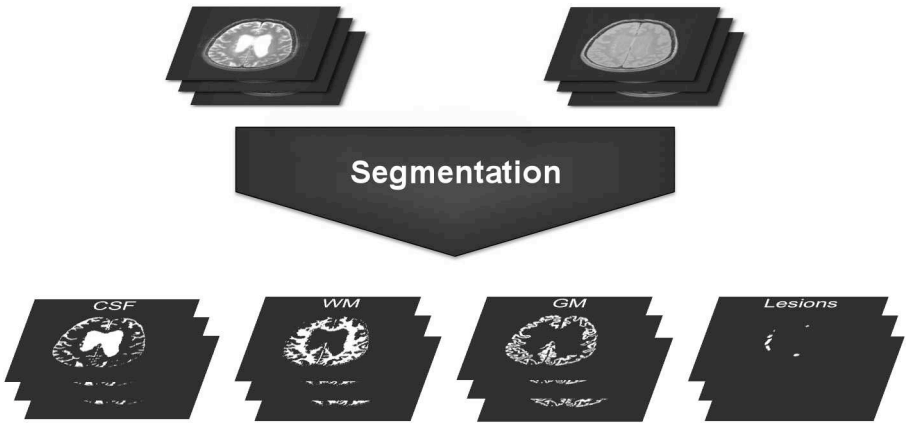


Figure 1.1: Tissue and lesion segmentation of two MRI sequences of the same patient.

Conventional magnetic resonance imaging techniques (MRI) are highly sensitive for detecting MS plaques and can provide quantitative assessment of inflammatory activity and lesion load. MRI-derived metrics have become established as the most important paraclinical tool for diagnosing MS, for understanding the natural history of the disease and for monitoring the efficacy of experimental treatments [56]. This image modality with its different sequences also offers a high contrast between the main brain tissues, namely the grey matter (GM), formed by neuron nuclei; the whitte matter (WM), formed by neuronal axons; and the cerebrospinal fluid (CSF) protecting the brain.

For quantitative analysis of focal lesions, in both individual and temporal studies, manual or semi-automated segmentations of different MRI sequences have been used to compute the total number of lesions and the total lesion volume (see figure 1.1). Manual delineation of MS lesions, however, is both challenging and time-consuming because of the large number of MRI slices for each patient which composes the three-dimensional information. Moreover, it is prone to intra-observer variability (the same study analysed by the same neuroradiologist at a different time) and inter-observer variability (the same study analysed by different neuroradiologists).

The development of fully automated MS segmentation methods, which can segment large amounts of MRI data and do not suffer from intra- and inter-observer variability, has become an active research field. Unfortunately, the results of these fully automated methods show less agreement with manually segmented scans than those obtained with segmentations of independent observers due to the intensity overlap with other features. Moreover, when evaluating MS segmentation methods still exists the lack of an in vivo

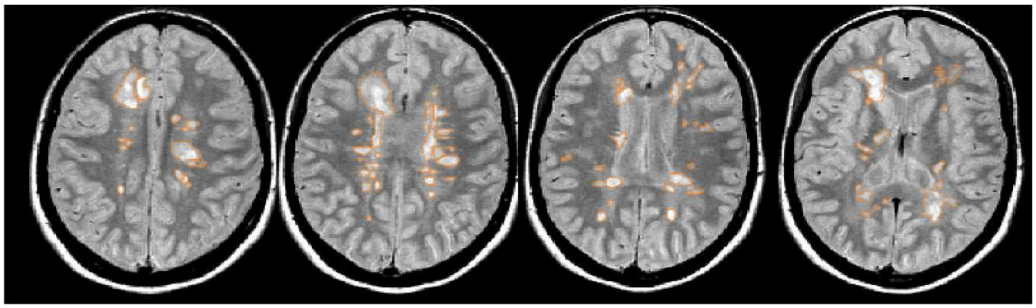


Figure 1.2: Lesion segmentation example. MS lesions have a small size and their intensity usually overlaps with other tissues.

method to obtain a reliable ground truth data mainly because of the large intra- and inter-observer variability. Notice that disagreements between segmentations may have large influences on some evaluation measures due to the small volume of the lesions (see figure 1.2).

1.2 Research framework

The research framework of this master thesis is located within two funded research projects (2009): "SALEM: Segmentación Automática de Lesiones de Esclerosis Múltiple en imágenes de resonancia magnética" (reference PI09/91918) awarded by the Instituto Carlos III, and the "Salem: toolkit para la segmentación automática de lesiones de esclerosis múltiple en resonancia magnétic" (reference VALTEC09-1-0025) by the Generalitat de Catalunya.

In both projects we are collaborating with relevant hospitals and medical expert teams in the field of multiple sclerosis such as the Hospital Vall d'Hebron, the Clínica Girona and the Hospital Dr. Josep Trueta. These 3 hospitals will provide the data from real patients, as well as manual expert annotations that will be used to test and evaluate the developed algorithms.

1.3 Objectives

In this master thesis **we investigate automatic methods capable of classifying brain tissues and detecting MS lesions in magnetic resonance imaging** usually aiding lesion detection with a previous tissue classification. The main focus is to use these

results to help in the diagnosis of the disease as well as its follow-up during drug therapy treatment.

The general goal of this master thesis can be subdivided in the following objectives:

- To analyse different tools for MRI preprocessing. These preprocessing steps can be divided in three main groups: noise reduction due to the capture process, the correction of the bias field inherent to this image modality and skull stripping to remove non-brain tissue that can bias segmentation results. Public software solutions will be tested as part of the whole preprocessing pipeline.
- To exhaustively analyse the state of the art of MS lesion segmentation techniques. This objective aims to review the whole MS lesion segmentation state of the art to understand better their advantages and drawbacks.
- To implement the techniques with the best reported results and test them with synthetic and real data to extract conclusions and point out strengths and weaknesses.
- To establish the key points of our proposal for MS lesion segmentation. This proposal will focus on improving weaknesses on the current state of the art, taking advantage on the strengths of the analysed techniques.

1.4 Planning

According to the Gantt diagram of figure 1.3 the master thesis will be developed in different tasks according to the objectives. Those tasks are summarised as follows:

- **Analysis of the state of the art on multiple sclerosis MRI segmentation.** During the first stage of this thesis, we will study the current state of the art on multiple sclerosis lesion segmentation of brain MRI. This will include the analysis of supervised segmentation strategies based on atlas and registration and publicly available preprocessing tools.
- **Analysis and implementation of the best techniques.** After analysing the state of the art, the best techniques according to their reported results will be selected in order to further study their advantages and drawbacks.
- **New technique proposal.** After an extensive analysis on all the approaches, we will implement a pipeline approach which will be tested and evaluated with both

synthetic and real data and we will define new research directions. Based on the weaknesses of the implemented methods, we will propose improvements.

- **Documentation.** All details regarding those steps will be documented in this master thesis memory.

1.5 Document structure

- **1 Introduction.** This chapter presents a brief summary of the background, objective and planning of this master thesis project. On the following chapters all details regarding to MS segmentation are presented and extended in order to present the current day techniques in this field and to introduce a new proposal.

- **2 Problem definition.** After presenting the problem in chapter 1, we will further define its background according to our research. The main topics described in this chapter are how to screen MS using MRI, which sequences are used for this purpose, why are they used for this purpose and which issues arise from the capture process according to the image processing viewpoint.

- **3 State-of-the-art.** A revision on the most recent techniques dealing with this problem is presented in this chapter, focusing on advantages and drawbacks. A classification of those techniques in supervised and unsupervised strategies is also introduced, emphasising on atlas-based segmentation methods. Finally, the reported results are gathered, as well as the most common evaluation measures and databases followed by the extracted conclusions.

- **4 Proposal.** The most significant techniques according to the reported results are analysed thoroughly in this chapter. Rigid and nonrigid atlas registration methods and publicly available preprocessing tools for brain MRI are also described. Based on our implementations a new proposal extending the framework is also introduced.

- **5 Results.** The implemented methods will be tested and evaluated with both synthetic and real data using common similarity measures. In this chapter we present our results pointing out strenghts and weaknesses.

- **6 Conclusions and future work.** In this final chapter, conclusions summarising the developed work are presented. Based on these conclusions, possible solutions are

also introduced to be implemented as future works on the framework of the PhD thesis.

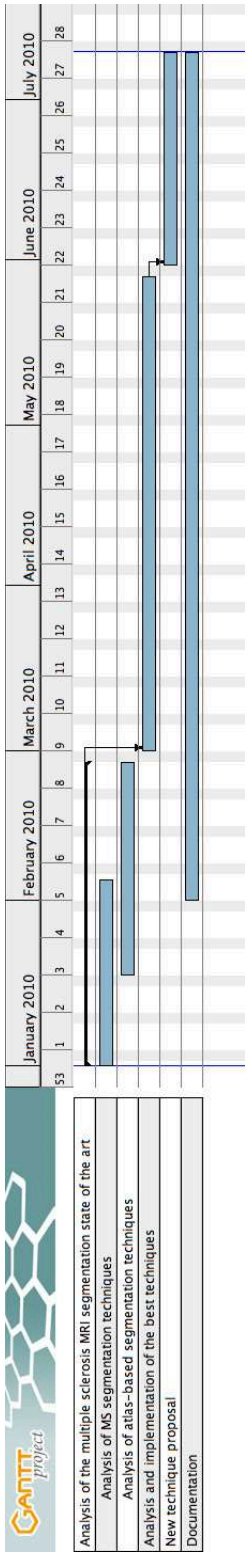


Figure 1.3: Gantt diagram representing the master thesis planning.

Chapter 2

Problem definition

Imaging is usually prefered over biopsy on clinical practice for several reasons. For instance, when using non-invansive techniques, collateral risks from operation procedures can be avoided. During the last years, different image modalities have appeared arising the question of what imaging technique should we use for MS diagnosis and follow-up and why should we use it.

2.1 Why MRI?

Magnetic resonance imaging is the most sensitive technique for the detection of demyelinating lesions on the central nervous system (CNS) in MS patients [34]. As a consequence of this high sensitivity, MRI has become an essential technique, not only on MS diagnosis, but also as a pronostic marker in the initial phase of the disease according to the frequency and gravity of future clinical recurrence, as well as future impairment rate [17, 49]. Moreover, MRI contributes to a better comprehension of its natural history and the effectiveness evaluation of new treatments [57, 36].

The new diagnostic criteria proposed by McDonald et al. [55] highlight MRI discoveries, by allowing to establish a MS diagnostic on patients with a single clinical episode when MRI proves the presence of spatially and temporally disseminated demyelinating lesions on the central nervous system [63].

2.2 What is MRI?

MRI systems, as shown in figure 2.1, consist of the following components:

- A large magnet to generate the magnetic field and shim coils to make the magnetic field as homogeneous as possible. This magnetic field aligns the hydrogen nuclei of the brain.
- A radiofrequency (RF) coil to transmit a radio signal into the body part being imaged. This radio signal is applied after aligning the hydrogen nuclei with the high magnetic field.
- A receiver coil to detect the returning radio signals due to the nuclei relaxation.
- Gradient coils to provide spatial localisation of the signals by first selecting a slice of interest and then selecting each location (according to a row and column). Those spatial locations comprise, commonly for MS, an anisotropic volume of $1 \times 1 \times (3 - 5)$ mm dimensions.
- A computer to reconstruct the radio signals into the final image (usually by means of Fourier transforms).

The voxel intensity on the MR image is determined by four basic parameters: proton density, T1 relaxation time, T2 relaxation time, and flow. Proton density is the concentration of protons in the tissue in the form of water and macromolecules (proteins, fat, etc). The T1 and T2 relaxation times define the way that the protons revert back to their resting states after the initial RF pulse. The most common effect of flow is loss of signal from rapidly flowing arterial blood.

The contrast on the MR image can be manipulated by changing the pulse sequence parameters. A pulse sequence sets the specific number, strength, and timing of the RF and gradient pulses. The two most important parameters are the repetition time (TR) and the echo time (TE). The TR is the time between consecutive 90 degree RF pulse. The TE is the time between the initial 90 degree RF pulse and the echo.

The most common pulse sequences are the T1-w and T2-w spin-echo sequences. The T1-w sequence uses a short TR and short TE ($TR < 1000$ ms, $TE > 30$ ms). The T2-w sequence uses a long TR and long TE ($TR > 2000$ ms, $TE > 80$ ms). The T2-weighted sequence can be employed as a dual echo sequence. The first or shorter echo ($TE < 30$ ms) is proton density (PD) weighted or a mixture of T1 and T2. This image is very helpful for

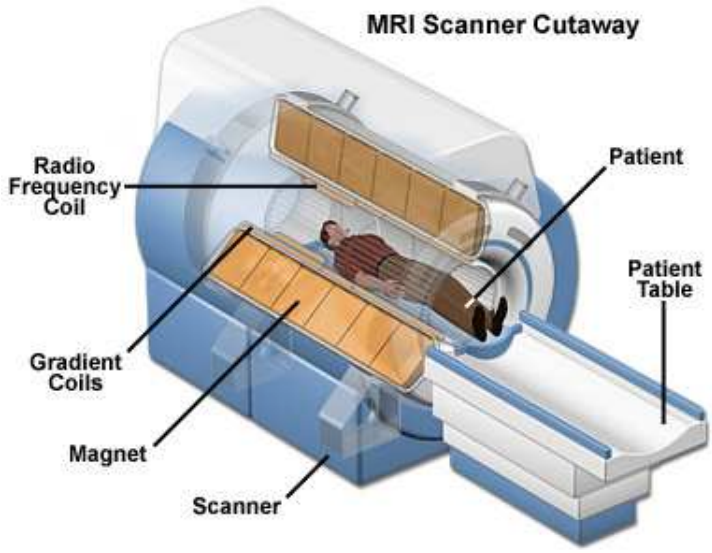


Figure 2.1: MRI scanner scheme with its main parts. Image extracted from http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri/.

evaluating periventricular pathology, such as multiple sclerosis, because the hyperintense plaques are contrasted against the lower signal CSF. More recently, the FLAIR (Fluid Attenuated Inversion Recovery) sequence has been introduced. FLAIR images are T2-w with the CSF signal suppressed.

However, the correlation between the burden of lesions observed on conventional MRI scans and the clinical manifestations of the disease remains weak. The discrepancy between clinical and conventional MRI findings in MS is explained, at least partially, by the limited ability of conventional MRI to characterise and quantify the heterogeneous features of MS pathology. Other quantitative MR-based techniques, however, have the potential to overcome such a limitation of conventional MRI. Indeed, magnetisation transfer MRI, diffusion tensor MRI (DTI), proton MR spectroscopy, and functional MRI (fMRI) are nowadays contributing to elucidate the mechanisms that underlie injury, repair, and functional adaptation in patients with MS [35, 54].

2.3 How are MR images of MS patients?

Looking at figure 2.2, contrast differences between conventional MRI caused by different relaxation times are clear. For instance, the CSF appears dark in both T1 and FLAIR,

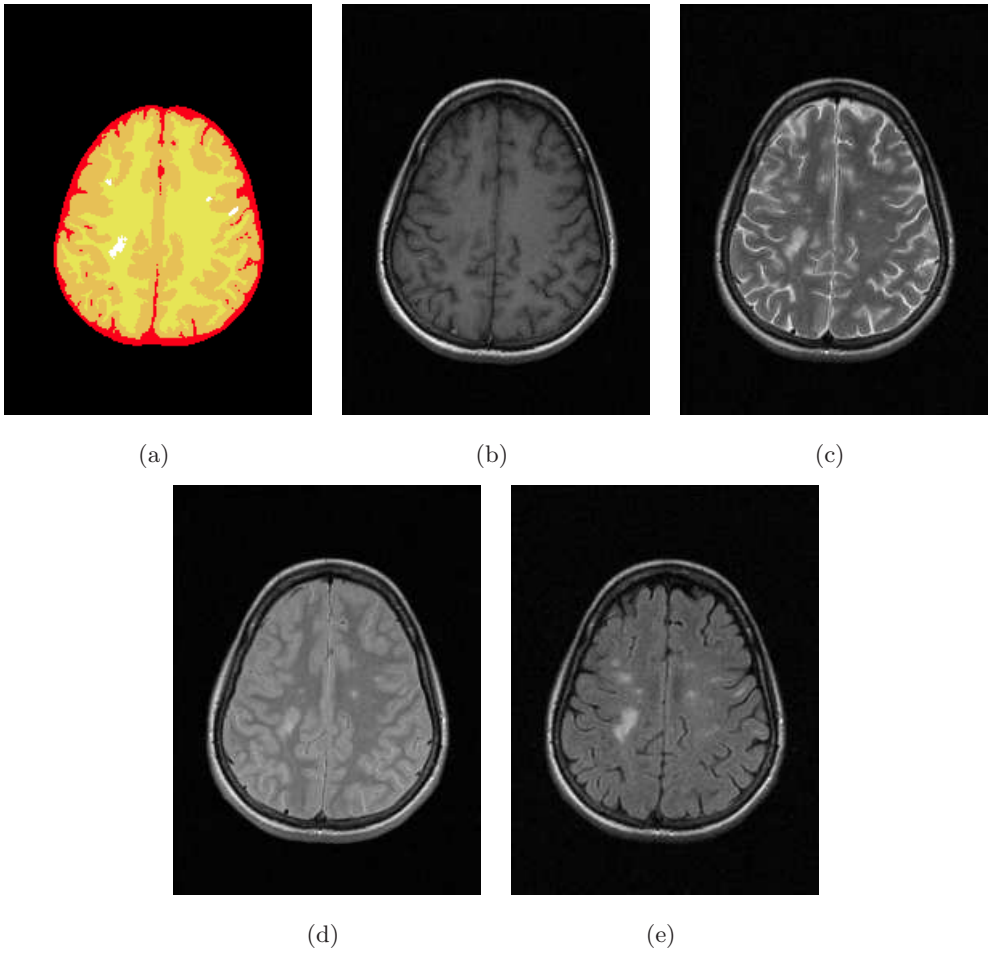


Figure 2.2: Different MRI sequences of the brain. a) Tissue segmentation: CSF appears red, GM appears orange, WM appears yellow and lesions appear white, b) T1 image, c) T2 image, d) PD image and d) FLAIR image.

while its the brightest tissue on T2 and has similar intensities to GM on PD. On the other hand WM is the brightest tissue on T1, has an intermediate grey level on FLAIR, similar to GM and has the lowest signal on both PD and T2. Finally, GM appears also with an intermediate grey level on T2 and T1 according to the other two brain tissues.

Both acute and chronic MS plaques appear as focal high signal intensity areas in T2-w sequences (including FLAIR), reflecting their increased tissue water content. The signal increase indicates edema, inflammation, demyelination, reactive gliosis and/or axonal loss in proportions that differ from lesion to lesion. They are typically discrete and focal at the early stages of the disease, but become confluent as the disease progresses.

Gadolinium-enhanced T1-w imaging (consisting on applying a contrast agent before aquiring the image) is highly sensitive in detecting inflammatory activity. This technique detects disease activity 5 to 10 times more frequently than clinical evaluation of relapses, suggesting that most of those **enhancing lesions (EL)** are clinically silent. Individual and temporal MRI studies have shown that the formation of new MS plaques is often associated with contrast enhancement, mainly in the acute and relapsing stages of the disease.

Approximately 10% to 20% of T2 **hyperintense lesions (HL)** are also visible on T1-w images as areas of low signal intensity compared with normal appearing white matter (WM). These so-called T1 **black holes (BH)** have a different pathological substrate that depends, in part, on the lesion age. Chronic black holes correlate pathologically with the most severe demyelination and axonal loss, indicating areas of irreversible tissue damage. T1-w sequences have a higher specificity than T2-w sequences for detecting lesions with irreversible tissue damage and may serve as surrogate markers of disability progression in clinical trials.

2.4 Is MRI perfect?

Even though MRI has become an interesting imaging tool, several issues arise from the capture process causing undesired image artifacts. Some of those artifacts are usually taken care of during the scanning procedure. For instance, motion artifacts due to the head movement can be corrected using fast sequences and constraining the patient's head. Some others may require a calibration procedure or rescan. For example, peak artifacts due to a high signal value that darkens the whole image or aliasing caused by using a small field of view.

However, some artifacts are inherent to the capture system and can not be dealt with during acquisition. As any capture system, signals are corrupted by noise, often assumed additive and following a Gaussian distribution. In the case of MRI this noise is caused by the gradient magnets (the ones responsible for voxel localisation). Inhomogeneities on the magnetic field and also patient properties cause a smooth inhomogeneity field across the MRI (see figure 2.3). This artifact is also known as bias field. Even though those two issues can not be avoided on acquisition, they can be reduced in order to increase the signal to noise ratio (SNR).

The gradient magnets are also responsible for the so-called partial volume effects (PVE).

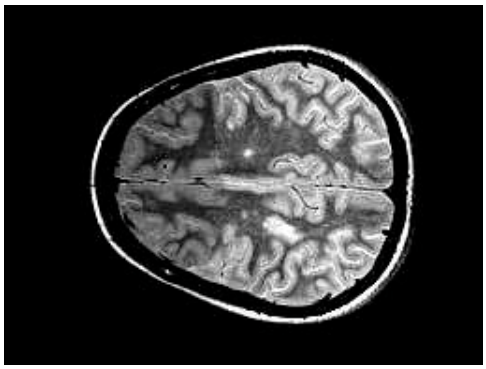
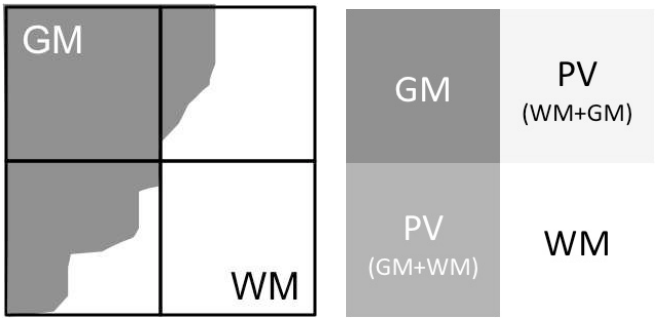


Figure 2.3: Slice of an MRI volume with clear intensity inhomogeneities (the leftmost part is darker than the righmost one) and noise (mainly on the WM).



(a) (b)

Figure 2.4: 2D synthetic example of the partial volume effects. a) Represents the real tissues and how they are sampled while b) shows the results of sampling zones with both tissues.

Localisation is done in order to scan a volume that may comprise more than one tissue. For instance, the cortex is folded with sulcus containing WM. Those frontier regions are usually scanned as a voxel, blurring the real edges between tissues as shown in figure 2.4.

Moreover, the presence of other nonbrain tissues (see figure 2.5) affects intensity distributions on the image. This is also inherent to the capture process but not as much an artifact as it is an issue for automatic segmentation systems. It is not clear how the probability density function of each main tissue (GM, WM and CSF) is altered by those external intensities, but segmentation results are usually improved when those voxels are masked out.

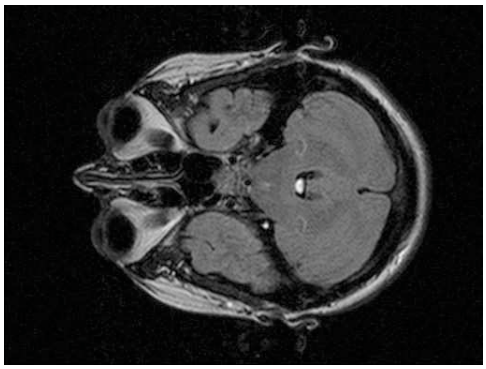


Figure 2.5: Slice of an MRI volume where different nonbrain structures appear (i.e. skull, eyes, fat).

2.5 What are we dealing with?

In this chapter we have seen why MRI has become a powerful technique on clinical practice for MS. Thanks to the presence on the brain of water molecules, and more precisely hydrogen nuclei, MRI scanners can provide volumetric soft tissue information with high contrast. Moreover, by tunning capture parameters, such as the pulse sequence, or relaxation times, different volume sequences can be acquired. The most widely used sequences on MS trials are PD-w, T1-w, T2-w and FLAIR images.

Focusing on MS lesions, those usually appear as bright spots on PD-w and T2-w images (including FLAIR sequences). Furthermore, all these lesions can be subdivided on 3 groups depending on the pathology and their properties on other sequences:

- Hyperintense lesions (HL): Those are the lesions that only appear on the abovementioned sequences and are the most frequent ones.
- Enhancing lesions (EL: These lesions appear as bright spots on T1-w images after applying a contrast agent, commonly Gadolinium. These lesions represent current imflammatory activity.
- Black holes (BH): These lesions are named by their presence, since they appear as "dark regions" on T1-w images. These lesions represent chronic and irreversible damage.

However, from the image processing viewpoint, MRI presents several issues that need to be taken into account. These issues include common Gaussian noise, a smooth bias field

causing brighter regions around the image, partial volume effects that can blur edges between tissues and the presence of nonbrain tissues that can bias the intensity distributions for each brain tissue.

Chapter 3

State-of-the-art

Automatic segmentation of MS lesions in brain MRI images has been widely investigated in recent years with the goal of improving MS diagnosis and patient follow-up. However, the performance of most of the algorithms still falls far below expert expectations. In this chapter, we review the main approaches to automated MS lesion segmentation, pointing out their strengths and weaknesses and suggesting new research directions. We classify the most recent important techniques into different strategies according to their main principle. We also present a qualitative and quantitative comparison of the results of the approaches analysed. Finally, we summarise the review and present a discussion on MS lesion segmentation.

3.1 Introduction

In this chapter, we review the state-of-the-art of strategies for automated MS lesion segmentation to point out their strengths and weaknesses and suggest new research directions. Moreover, we describe significant works in this field and classify the different techniques according to the strategy used. In particular, we first distinguish between supervised and unsupervised segmentation strategies, dividing the supervised group into atlas-based approaches and those based on training by means of features extracted from manual segmentations. We further divide the unsupervised group into those that use tissue segmentation to obtain the lesions and those that use only the lesion properties for the segmentation.

Various reviews of brain MRI segmentation have been presented in the past. For instance, Bezdek et al. [14] analysed 90 papers on MRI segmentation using pattern recognition techniques. The authors suggested dividing the algorithms into two categories:

supervised methods (such as Bayes classifiers with labelled maximum likelihood estimators, the k-Nearest Neighbour rule (kNN), and Artificial Neural Networks (ANN)) and unsupervised methods (i.e. Bayes classifiers with unlabelled maximum likelihood estimators or the fuzzy c-means (FCM) algorithms). In addition Clarke et al. [20] reviewed not only methods for MRI segmentation, but also general pre-processing algorithms, validation methods and registration between different MRI images.

However, these reviews were only related to soft brain tissue segmentation. More recently, Souplet et al. [73] presented a review of semi-automated and automated MS lesion segmentation approaches, analysing MS lesions, pre-processing steps and segmentation approaches. However, to the best of our knowledge, this chapter is the first attempt to review the most relevant works in automated MS lesion segmentation that provides an evaluation of the experimental results.

3.2 Classification of lesion segmentation approaches

In this section, we describe the MS lesion segmentation methods according to our classification shown in tables 3.1 and 3.2. Notice that these tables offer an overview of all these works with respect to the strategy, the type of MRI images used, the type of lesions detected and the algorithms. The criteria used to select these methods are based on several aspects: 1) representative works for each of the identified strategies; 2) the reported experimental results and the evaluation measures used by the authors; and 3) the data sets used to perform the experiments (synthetic, real cases and data from the MS Lesion Segmentation Challenge 2008 [75], which enables a quantitative comparison of the evaluation results).

We can distinguish between supervised and unsupervised segmentation strategies. The supervised approaches described in table 3.1 are those based on using some kind of a priori information or knowledge to perform the MS lesion segmentation. Within the group of supervised strategies, we identify two sub-groups of approaches. In the first group all the approaches use atlas information and therefore require the application of a registration process to the analysed image to perform the segmentation. In the second group all the approaches perform an initial training step on features extracted from manually segmented images annotated by neuroradiologists. The systems in this second group employ the image intensities previously segmented by an expert to train a classifier that segments the tissues and lesions of the MRI images.

Table 3.1: Summary of the examined supervised MS lesion segmentation approaches with respect to the sequences, algorithms and lesions. The approaches are classified according to the strategies. The acronyms for the algorithms stand for (in alphabetical order): Artificial Neural Networks (ANN), Expectation Maximisation (EM), Fast Trimmed Likelihood Estimator (FAST-TLE), Fuzzy C-Means (FCM), Fisher Linear Discriminant (FLD), Hidden Markov Chains (HMC), K-Nearest Neighbours (kNN), Mean Shift (MeS), Morphological Greyscale Reconstruction (MGR), Markov Random Fields (MRF), Principal Components Analysis (PCA). The acronyms for the lesions and sequences stand for: Diffusion Tensor Imaging (DTI), Fractional Anisotropy (FA), Mean Diffusivity (MD).

		Article	Algorithms	Images	Lesions
Supervised	Atlas	Van Leemput (2001) [53]	EM + MRF	PD, T1, T2	HL
		Zijdenbos (2002) [87]	ANN	PD, T1, T2	HL
		Wu (2006) [86]	kNN	PD, T1c, T2	EL & BH & HL
		Shiee (2008) [69]	FCM	PD, T1, T2, FLAIR	HL
		Shiee (2008) [68]	FCM	T1, T2, FLAIR	HL
		Bricq (2008) [18]	FAST-TLE + HMC	T2, FLAIR	HL
		Prastawa (2008) [61]	Region partitioning	T1, T2	HL
		Kroon (2008) [51]	PCA	T1, T2, FLAIR, DTI (FA, MD)	HL
		Souplet (2008) [72]	EM + GMM	T1, T2, FLAIR	HL
		Tomas (2009) [78]	Bayes	T1, T2, FLAIR	HL
		Akselrod-Ballin (2009) [1]	FLD + Decision Forest	PD, T1, T2, FLAIR	HL
		de Boer (2009) [28]	kNN	T1, T2, FLAIR	HL
		Shiee (2010) [70]	FCM	T1, T2, FLAIR	HL
	Manual segmentations	Kamber (1995) [48]	Different classifiers	PD, T1, T2	HL
		Goldberg (1998) [41]	ANN	PD, T2, FLAIR	EL & HL
		Alfano (2000) [2]	Spatial Clustering	PD, T1, T2	HL
		Anbeek (2004) [4]	kNN	PD, T1, T2, FLAIR, IR	HL
		Anbeek (2005) [3]	kNN	PD, T1, T2, FLAIR, IR	HL
		Sajja (2006) [65]	Parzen windows	PD, T2, FLAIR	HL
		Datta (2006) [27]	Parzen windows + MGR	PD, T1, T2, FLAIR	BH
		Anbeek (2008) [5]	kNN	FLAIR	HL
		Scully (2008) [66]	Bayes	T1, T2, FLAIR	HL
		Morra (2008) [58]	AdaBoost	T1, T2, FLAIR, DTI (FA, MD)	HL
		Subbanna (2009) [76]	Simulated annealing + MRF	PD, T1, T2	BH & HL
		Lecoeur (2009) [52]	Graph Cuts	PD, T1, T2	HL

With regard to table the unsupervised strategies in table 3.2, where no prior knowledge is used, we identify a sub-group of methods that segment brain tissue to help lesion segmentation and another sub-group that uses only the lesion properties for segmentation. In the first sub-group, there are methods that either segment the tissue first and then the MS lesions, or segment the tissue and the lesions at the same time. In the second sub-group, the methods directly segment the lesions according to their properties, without providing tissue segmentation. The advantage of segmenting the tissue is that neuroradiologists can also evaluate the GM tissue volumetry and monitor the progression of cerebral atrophy.

Analysing the literature, we can also distinguish between single-channel or multi-channel approaches, i.e. approaches that use only one MRI image or those that combine several

Table 3.2: Summary of the examined supervised MS lesion segmentation approaches with respect to the sequences, algorithms and lesions. The approaches are classified according to the strategies. The acronyms for the algorithms stand for (in alphabetical order): Adaptive Mixtures Method (AMM), Constrained Gaussian Mixture Models (CGMM), Expectation Maximisation (EM), Fuzzy C-Means (FCM), Mean Shift (MeS), Morphological Greyscale Reconstruction (MGR), Markov Random Fields (MRF). The acronyms for the lesions and sequences stand for: Attenuation of Fluid by Fast Inversion Recovery with Magnetisation Transfer Imaging with Variable Echoes (AFFIRMATIVE).

		Article	Algorithms	Images	Lesions
Unsupervised	Tissue	Freifeld (2007) [37]	EM + CGMM	PD, T1, T2	HL
		Khayati (2008) [50]	Bayes + AMM + MRF	FLAIR	HL
		García-Lorenzo (2008) [38]	EM + MeS	T1, T2, PD	HL
		García-Lorenzo (2008) [40]	EM	PD, T1, T2	HL
		García-Lorenzo (2008) [39]	EM	T1, T2, FLAIR	HL
	Lesion	Bedell (1998) [12]	Threshold Substraction	T1, T1c, AFFIRMATIVE	EL
		Bourdraa (2000) [16]	FCM	PD, T2	HL
		He (2002) [44]	MGR	T1, T1c, AFFIRMATIVE	EL
		Datta (2007) [26]	MGR	T1, T1c, T2, FLAIR	EL

images. Single-channel approaches are mainly used to segment the brain tissues. For instance, T1-w images are widely used for this purpose, since they show the best contrast between the three main brain tissues: WM, GM and CSF. This initial tissue segmentation may then be used to help obtain the final lesion segmentation, and T2-w and PD-w are the classical images for detecting MS lesions. Another example of the single-channel approach is the segmentation of MS lesions using just the FLAIR sequence [50]. The multi-channel approaches, on the other hand, use at least two of the PD-w, T1-w, T2-w and FLAIR images. One of the benefits of using more than one of the different MRI images is that it increases the intensity feature space, producing a better discrimination between brain tissues. Furthermore, more than one kind of image may be required because MS lesions can appear independently in different images [86], depending on their subtype. As shown in tables 3.1 and 3.2, most of the approaches combine different MRI images to perform both the tissue and lesion segmentation.

3.2.1 Supervised strategies based on atlas

If we look at the strategies based on atlas information, we can distinguish between the use of both statistical and topological atlases. A statistical atlas provides the prior probability of each voxel belonging to a particular tissue class. This statistical atlas is built from a set of manual segmentations of the structures of interest, where the boundaries of each structure are used to make a smooth probability map and to account for anatomical

variations beyond those present within the training set. Notice that the use of an atlas can be helpful in classifying tissues in the presence of noise or inhomogeneities (an atlas takes spatial information into account), or in order to segment lesions as deviations from normal human brains. On the other hand, a topological atlas is a parcellation of the brain that is edited to encode a specific topology for each structure and group of structures. This topological atlas is usually used to preserve topology and to lower the influence of competing intensity clusters in regions that are spatially disconnected, while the statistical atlas affects the segmentation of adjacent structures that have similar intensity.

Notice that in both the statistical and the topological atlas methods, the analysed MRI image has to be registered with the atlas before the segmentation is done as shown in figure 3.1. Hence, the challenge for these atlas-based approaches is to align the atlas and the images, thereby converting the segmentation problem into a registration problem. Moreover, segmentation and registration can be combined in a probabilistic framework via the Expectation-Maximisation algorithm (EM) [29] or the Fuzzy C-Means (FCM) [13]. The EM algorithm computes the probability to belong to a certain distribution for each entry of the data set, and afterwards estimates the hidden paramaters of this distribution that maximise the previous expectation in an iterative manner until convergence is reached. Even though great convergence properties are proven, this algorithm can lead to non-desired local optima, especially when only relying on the data itself. Therefore, proper initialisation, as well as spatial information are introduced into the framework using atlas-based approaches. On the other hand, FCM inherently treats all classes as Gaussians with the same variance, since it only takes class centroids and membership values for each voxel into account. Again, no spatial information is encoded in the original algorithm.

For instance, Van Leemput et al. [53] extended their previous work [79], based on the EM algorithm and bias field correction, by searching for outliers that follow a set of user defined rules (i.e. lesions should appear hyper-intense on both PD-w and T2-w images) and checking wether or not they are in the vicinity of WM and CSF (since around $90-95\%$ of multiple sclerosis lesions appear on WM). In a similar way, Bricq et al. [18] decided to apply their hidden Markov chain approach [19] to detect multiple sclerosis lesions as outliers while Shiee et al. [69, 68, 70] modified their fuzzy segmentation algorithm [11] to segment lesions inside WM structures.

In the same way, Souplet et al. [72] segment brain tissues on multiple sclerosis scans using the EM algorithm, previously initialised by an atlas. 11 binary images are obtained, namely: GM, WM, CSF, 6 partial volume classes and 2 outliers classes. Using the 3

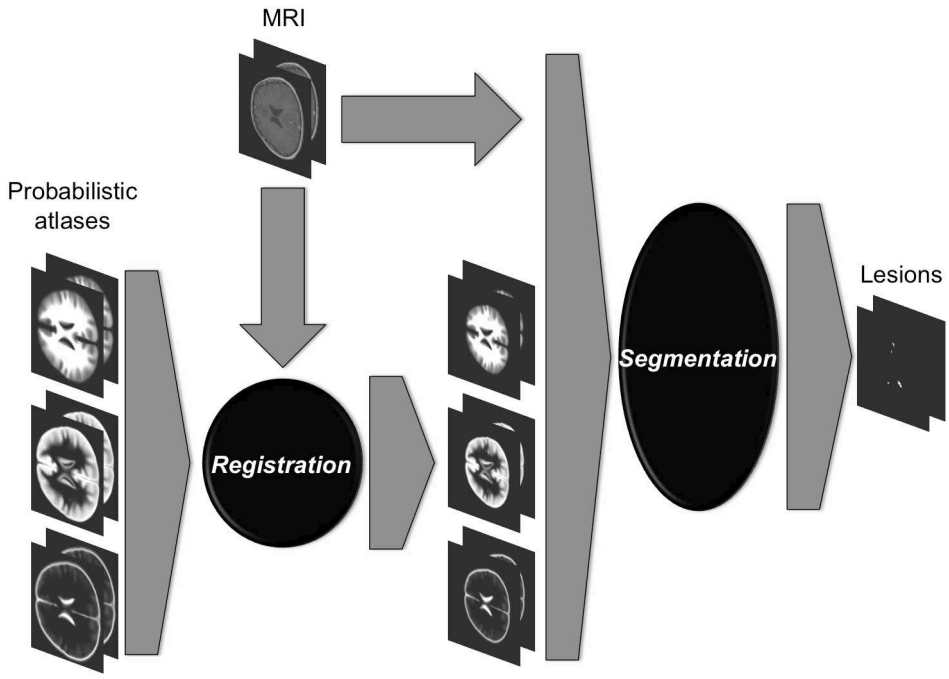


Figure 3.1: Atlas-based segmentation scheme. First the atlas is registered to the new image and afterwards the segmentation is performed.

pure tissue masks, normal appearing tissue parameters (mean and standard deviation) are learned on the T2-FLAIR image to define a lesion threshold. Afterwards, those initial lesions are refined by applying WM masks (given by the EM algorithm).

Those approaches present a major drawback since parameter estimation can be biased by unhealthy tissue types. Thus, supervised methods relying on tissue samples from lesion and non-lesion classes should perform better, since no intensity distribution or model is assumed. For instance, Kamber et al. [48] compared three different classifiers: a minimum distance classifier (which assigns the label of the closest mean class), a Bayesian classifier (based on the Bayes theorem) and a decision tree (ID3). In their work, a probabilistic atlas was used to provide features to the classifiers, as well as to constrain the search on WM masks. Those two ideas have been used later on to segment and detect white matter lesions. The k-nearest neighbours (kNN) [32] is conceptually the easiest of supervised classification methods in its naïve implementation. This algorithm looks for the minimum distance between all the training examples and the one being tested and then assigns the most voted class among the closest neighbours (samples with minimum distances). However, kNN makes strong assumptions about the data, i.e. that there is no correlation

among different multivariate channels and that all variances are the same Based on this classifier, Wu et al. [86] implemented a kNN classifier trained with 20 pixels for each class. After classification, a probabilistic atlas was used to relabel GM and multiple sclerosis lesions applying WM masks.

On the other hand, Zijdenbos et al. [87] used a probabilistic atlas as features to an artificial neural networks (ANN) method. The input of this well-known classification technique was based on three MRI modalities (PD-w, T1-w and T2-w) and three spatial tissue priors (WM, GM and CSF) coming from the atlas. Moreover, healthy atlases could be extended by applying lesion probabilities, enhancing their power as input features for a classification technique. In this way, Kroon et al. [51] decided to manually warp segmented lesions to a publicly available atlas and apply this new atlas as a feature to a PCA-based classification framework. This algorithm is trained with lesion and non-lesion samples. On a more general framework, Akselrod-Ballin et al. [1] proposed to segment the volume in different regions using a graph based algorithm. Those regions are then characterised with a rich set of extracted features (comprising probabilities coming from an atlas) and classified using a decision forest along with the Fisher linear discriminant. This combination of segmentation and region classification helps reducing misclassification at voxel level owing to noisy data.

Following the tissue segmentation approach of Cocosco et al. [21], atlases can also be used to select healthy tissue training points for a classifier algorithm. For instance, de Boer et al. [28] extended healthy tissue classification using a kNN algorithm with white matter lesion segmentation. Upon completion of the firt step in which CSF, GM and WM are segmented, lesions appear as GM with a "halo" of WM. Afterwards, a histogram of all GM voxels is created on this volume and a threshold is defined to segment the lesions.

Furthermore, atlases also provide a way to select abnormal tissue samples while estimating healthy ones. For example, Prastawa et al. [61] proposed to use the Minimum Covariance Determinant to estimate tissue PDFs using healthy samples. Outliers that follow a set of rules (such as lesions appear bright in FLAIR images) are considered multiple sclerosis samples. After PDF estimation, the volume is partitioned using the watershed transform. In a final step, those regions are classified using the previously computed PDFs.

Finally, a combination of training sample points and WM mask refinement for multiple sclerosis segmentation was presented by Tomas and Warfield [78]. This approach used both a set of topological atlases for defining healthy tissue samples by combining them using the STAPLE algorithm [85] and a statistical atlas resulting from averaging those

manual segmentations. Afterwards, multiple sclerosis samples were defined as intensity outliers by comparing the reference group and the subject volumes. Subsequently, a Bayes classifier was trained to select lesion and non-lesion voxels. Since some of those voxels were misclasified as false positives, this classification was refined using WM masks extracted from the statistical atlas.

In conclusion, we want to mention that these atlas-based approaches can be used to segment both the tissue and the lesions. Moreover, atlases make it possible to treat the lesions as outliers in the tissue, to introduce spatial information into the segmentation process and to reduce the false positive lesion segmentations. As a drawback, these approaches rely on building an atlas, which is not a simple task. In addition, they also introduce the registration problem into the MS lesion segmentation. Note that this registration step is even more difficult when dealing with cases with severe atrophy, large numbers of lesions, etc.

3.2.2 Supervised strategies based on learning from manual segmentation

The second group of supervised approaches uses manually-segmented images annotated by neuroradiologists to segment the MS lesions. However, unlike atlas-based approaches, which require a registration process between the analysed images and the atlas, these methods mainly use the image intensities of different MRI images to train a classifier for the segmentation purpose as shown in figure 3.2. As reported by several authors, the use of prior knowledge to guide the segmentation of MS lesions improves the robustness of the algorithms, thus reducing the volume of false positive lesions compared to purely data-driven segmentations.

We want to stress that some of the approaches classified in this category may be similar to strategies described in the previous section. However, the majority of the strategies included here rely on a training process performed using features extracted from manually-segmented MRI images. Furthermore, some of these methods include the use of registration algorithms that focus on the intra-sequence and inter-sequence pre-processing registration steps. Table 3.1 shows that a large number of proposals have followed this strategy, most of them being multi-channel approaches. Furthermore, different classifiers or a combination of them, for example, ANN, kNN, AdaBoost, Bayesian classifiers or decision trees, have been used to perform the segmentation.

The first example in table 3.1 is the proposal by Kamber et al. [48], where the inputs for the training step are not voxel intensities but rather their probabilities of belonging

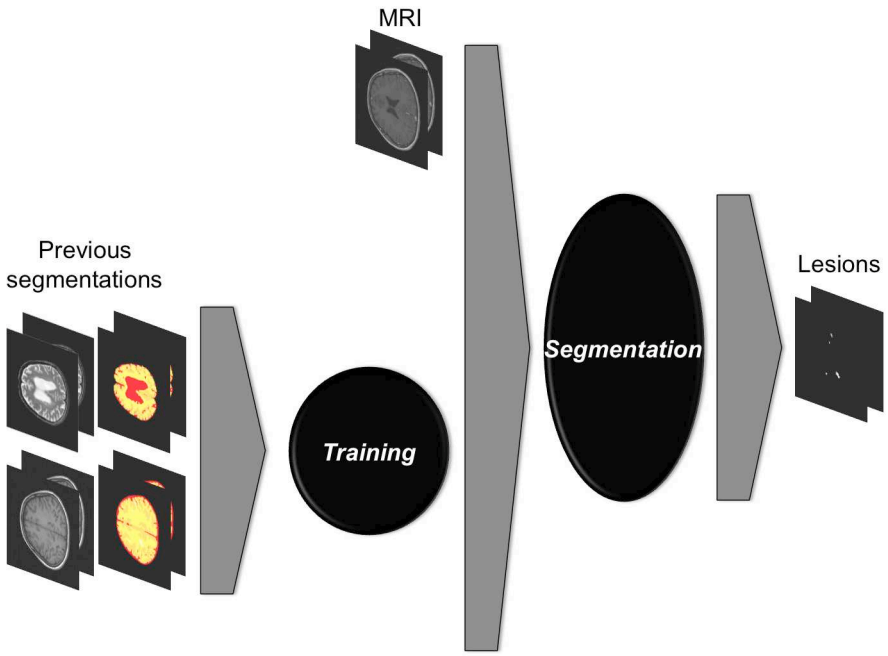


Figure 3.2: Segmentation using manual annotations scheme. First the system is trained - either by learning intensity tissue models or by creating a classifier - using previously segmented images and afterwards the segmentation is performed.

to WM, GM, or CSF tissue categories. Using this prior information as a pattern, their approach trains and tests a set of different classifiers to segment MS lesions. Goldberg et al. [41] use local thresholding to select the brightest regions of the image. Afterwards, the lesions are segmented by looking for closed contours and using different morphological properties such as area, perimeter and shape. For segmentation, an ANN is trained and used to classify the regions. Alfano et al. [2] also used previous segmentations of normal tissues to extract features used to train a spatial clustering algorithm. The authors stated in their experiments that their approach was also suitable for monitoring changes in the disease over time.

Anbeek et al. applied the kNN rule in two different studies. The aim in the first was simply to detect MS lesions [4] and spatial features were included in the classifier to achieve better lesion segmentations; in the second study, the aim was to model all brain tissues [3] and in this case, the kNN classifier was used to classify tissues and lesions simultaneously. It is important to mention that the authors tested their multichannel approach using information from T1-w, inversion recovery (IR), PD-w, T2-w and FLAIR

images, concluding that the incorporation of the T1-w, PD-w or T2-w did not significantly improve the segmentation results of the different brain tissue types.

From a different viewpoint, Sajja et al. [65] proposed segmenting CSF and lesions using a Parzen windows classifier and then segmenting WM and GM using a parametric method. Their assumption behind this approach is that GM and WM, but not lesions and CSF, follow a Gaussian distribution. Therefore, the authors first classify CSF and hyperintense T2-w lesions using a Parzen classifier and then the remaining brain parenchyma - excluding CSF and lesions - is classified into GM and WM using the PD-w and T2-w images and a MRF together with an EM algorithm. This method also exploits contextual information by using a fuzzy-connectedness to minimise the false negative lesion classifications. In another work, the authors also proposed segmenting black holes (considered as regional minima) in MS using a similar strategy [27]. After applying the previous segmentation method, black hole segmentation is achieved by using morphological greyscale reconstruction (MGR) in T1-w images.

The work of Scully et al. [66] introduced a new parametric method to the field of MS lesion segmentation. This method uses a vector image joint histogram, built over a training set, as an explicit model of the feature vectors indicating lesion. This model is then used to generate samples to train a naive Bayesian classifier which proceeds to classify the vector image composed of the T1-w, T2-w and FLAIR images. Using a different strategy, Morra et al. [58] proposed a framework to automatically segment sub-cortical structures in brain MRI images. Their method uses an AdaBoost algorithm to learn a unified appearance and context model which is then used to perform the lesion segmentation. Their feature pool includes intensity, position and neighbourhood features.

Subbanna et al. [76] presented a fully automated framework for identifying MS lesions in multi-channel MRI images. Manual segmented images are used to extract intensity histograms of both tissue and lesions. From the histograms, multivariate Gaussian distributions are estimated and used in the MRF classification step, which incorporate local spatial variations and neighbourhood information. Finally, Lecoeur et al. [52] also presented an optimised supervised lesion segmentation method using multi-channel MRI images. Their proposal creates an optimised spectral gradient colour space from single-channel images. Based on this transformation, they then apply graph cuts segmentation. As argued by the authors, the graph cuts algorithm provides an optimal solution for the joint use of regional and border information in a way that is similar to how MRFs work.

To summarise, the use of manually annotated data allows expert knowledge to be in-

corporated into MS segmentation approaches. Moreover, as shown in table 3.1, having initial segmentations allows a large variety of classifiers (ANN, kNN, EM, Bayes, etc.) to be applied. As in the atlas-based approaches, selecting a good initial MRI training set is an important step. Another issue with many lesion segmentation algorithms, particularly those that employ training data to model lesion intensity profiles, is that they are dependent on a specific acquisition sequence. These approaches must be modified or re-trained to process data acquired using alternative pulse sequences.

3.2.3 Unsupervised strategies segmenting tissue

There are several works in which the main principle consists of applying an unsupervised algorithm to segment the tissue and the lesions as shown in figure 3.3. These approaches detect lesions as outliers on each tissue rather than adding a new class to the classification problem. For instance, Freifeld et al. [37] first initialise their algorithm based on a pre-segmentation using K-means and its subsequent decomposition into a mixture of many spatially-oriented Gaussians per tissue (constrained GMM) in order to capture the spatial layout [42]. The intensity is considered as a global parameter and is constrained to be the same value for a set of related Gaussians per tissue. In order to detect the lesions, a set of rules that distinguishes between normal tissue regions and lesions is defined. Following initialisation, voxel-wise GMM parameters are learned via an EM algorithm. Finally, an active contour algorithm is used to delineate lesion boundaries.

García-Lorenzo et al. [38] combined a modified version of the EM-based method (mEM) to the trimmed likelihood estimator with the mean-shift algorithm [23] to segment MS lesions. A local segmentation approach (mean-shift) is used to generate local regions in the images, while an EM variant is employed to classify the regions obtained into healthy tissue or lesions. In another work by García-Lorenzo et al. [40, 39], the authors presented a modified version of the Spatio Temporal Robust Expectation Maximisation [9] to perform the MS lesion segmentation. Their approach is based on three main processes: robust estimation of healthy tissues using the mEM introduced in [38] (Gaussian mixture model is assumed), refinement of outlier detection and application of lesion rules. Khayati et al. [50] combined an adaptive mixture method, MRF and a Bayesian classifier to simultaneously classify the three main brain tissues and the MS lesions using only FLAIR images. In particular, they first propose to segment the brain into four classes: WM, GM, CSF and "others" Afterwards, inside the "others" class, lesions are dealt with as outliers not correctly explained by the model.

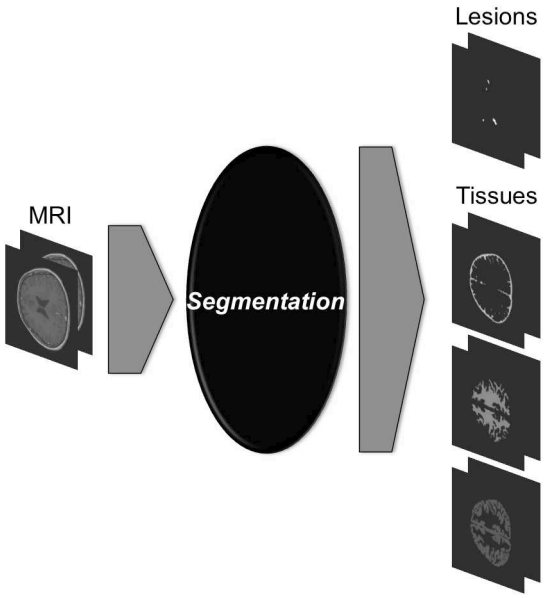


Figure 3.3: Tissue and lesion segmentation scheme. Those methods benefit from the tissue classification to perform the segmentation of MS lesions.

In conclusion, we have presented a set of algorithms that classify the outliers coming from a previous tissue segmentation in order to provide the lesion segmentation. Notice that the results of the lesion segmentation depend highly on the quality of the tissue segmentation. Furthermore, not all the tissue segmentation methods take abnormal cases (with severe lesions, atrophy, etc.) into account.

3.2.4 Unsupervised strategies segmenting only lesions

This last group of unsupervised approaches is based on using only the lesion properties to perform the MS lesion segmentation as shown in figure 3.4. Examples of this strategy include studies by Bedell and Narayana [12] and Boudraa et al. [16]. Bedell and Narayana [12] presented an automated segmentation and quantification of contrast-enhanced lesions based on performing threshold subtraction to eliminate enhancing structures such as choroid plexus. The authors reported that all MS lesions larger than $5mm^3$ were successfully identified and the automated analysis produced no false positive or false negative lesions above this volume in 13 different patients. The method used by Boudraa et al. [16] performs a FCM algorithm two times to detect lesions. The first FCM algorithm has the goal of obtaining two clusters: one that groups together CSF and lesions,

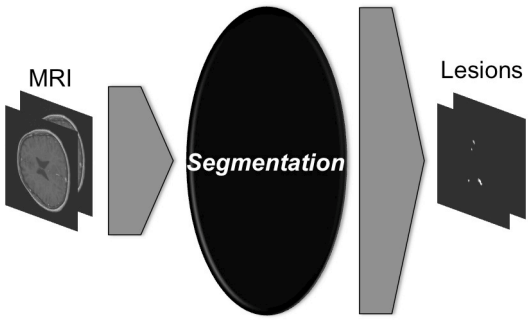


Figure 3.4: Lesion segmentation scheme. All those approaches rely solely on image itensities and lesion properties.

and another that groups together WM and GM. Afterwards, a second FCM algorithm is applied to distinguish between lesions and CSF. A final post-processing step based on anatomical knowledge is performed to remove extra segmented structures. This strategy assumes that, in the first step, all the lesions have been grouped together within a specific cluster and then, in the second step, this cluster is resegmented to distinguish between healthy tissue and lesions, taking spatial information into account.

He and Narayana [44] also proposed a method for the automated identification and segmentation of contrast-enhanced MS lesions in brain MRI images. This method relies on an adaptive local segmentation derived from morphological grey-scale reconstruction operations to identify both lesion and non-lesion enhancements. Similarly, Datta et al. [26] developed a method for the identification and quantification of gadolinium enhancements. This is also a multi-channel MRI approach and aims to identify enhancements using morphological operations. These enhancing lesions are further segmented based on fuzzy connectedness.

In regard to this strategy, we want to stress that, unlike the methods presented in the previous section, these approaches do not rely on an initial WM, GM and CSF tissue segmentation. Moreover, we have seen that they are useful for segmenting special lesions such as black holes and enhancing lesions. As a drawback, however, these approaches have to specifically define the properties used in each image, which is not an easy task. It should be noted that some artefacts may share the same lesion properties.

3.3 Reported results

MS lesion segmentation approaches are usually evaluated using different quantitative measures and both synthetic and real MRI volumes. In this section, we first present the most common data sets used in the works we have analysed and then describe the typical measures computed for the evaluation. Finally, we compare and discuss the results presented by the different approaches, highlighting the most interesting aspects.

3.3.1 MS databases

One of the main difficulties when comparing MS lesion segmentation approaches is the lack of a common database (with ground truth data) for training and testing the algorithms. Fortunately, this is becoming less of an issue with the appearance of new databases designed to meet this particular objective.

Synthetic databases are useful as an initial framework for testing the various approaches to segmentation. BrainWeb [22] is becoming a standard synthetic database for evaluating both tissue classification and MS lesion segmentation algorithms. BrainWeb contains simulated T1-w, T2-w and PD-w brain MRI data based on two anatomical models: 1) a normal brain and 2) a brain containing MS lesions. Moreover, this database provides different models according to parameters such as slice thicknesses, noise levels and levels of non-uniformity intensity. These data sets are available in three orthogonal views - axial, sagittal and coronal as shown in figure 3.5 - although the majority of algorithms use only the axial view.

Even though synthetic data sets are useful for an initial evaluation, they are not as challenging as real data sets. In most cases, algorithms that are correctly tuned for synthetic data may not be successfully applied to a real in vivo acquired volume. Therefore, segmentation algorithms have to be tuned and tested with MRI volumes of real patients. A review of the work of recent years has highlighted a noticeable lack of a public MS MRI databases supplied by hospitals. However, the recent MS Lesion Segmentation Challenge (2008) [75] has provided a common framework for evaluating these algorithms, and for making comparisons among them. The MRI data for this competition was acquired separately by the Children's Hospital Boston (CHB) and the University of North Carolina (UNC). In total, 53 brain MRI volumes hat were randomised into three different groups: 20 training MRI volumes (including ground truth); 25 testing volumes (without ground truth) that were downloadable before the contest for training/validating the different al-

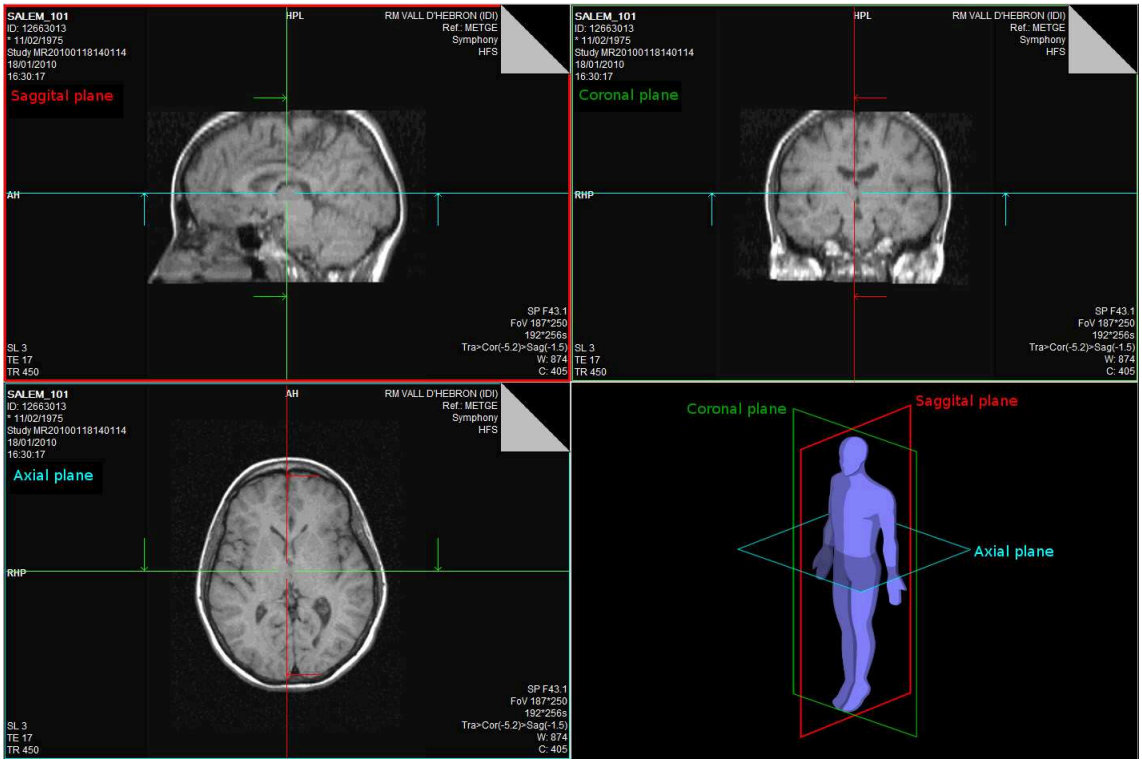


Figure 3.5: 3D planes scheme.

gorithms; and eight testing volumes used for the real contest from which the different participants extracted their on-site results.

3.3.2 Evaluation measures

The results are evaluated in different ways in the reviewed papers. However, all the measures are based on comparing the result of automated segmentations with the ground truth, which is usually annotated by an expert. In order to avoid intra- and inter-observer variability, segmentations from more than one expert should be used, since this provides a more consistent ground truth. Strategies such as the STAPLE algorithm [85] allows annotations from different experts to be fused. This is an important issue since, bearing in mind the small volume of each lesion, differences in the evaluation measures can be significant.

Automated segmentations and ground truth can be compared by either comparing each voxel in each lesion (voxel-to-voxel), or using the whole detected lesion (lesion-to-lesion). Note that in both cases, voxels and lesions can be classified as a true positive (TP), false

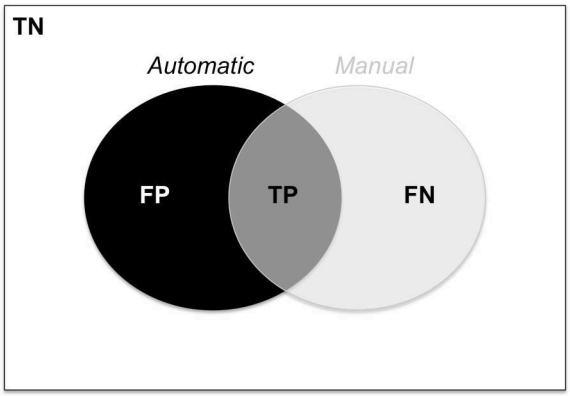


Figure 3.6: Comparison between manual and automated segmentation. TP represent those lesions where both automatic and manual segmentations, while TN represent those zones without lesions where both segmentations agree. According to disagreements, FP are those lesions segmented only by automated methods and FN are those lesions segmented manually.

positive (FP), true negative (TN) or false negative (FN), as shown in figure 3.6. However, a criterion must be defined when using lesion-to-lesion evaluation. Obviously, the objective is to obtain the maximum number of TPs and TNs, and at the same time reduce the number of FPs and FNs. However, in practice, one should find the best trade-off between these values, since increasing the number of TPs usually increases the number of FPs, while reducing the number of FNs also reduces the number of TNs. In fact, there is permanent debate about the best option: should we obtain more TPs or reduce FNs? One could argue that it is preferable to reduce the FNs at the expense of increasing FPs. However, increasing the number of FPs also leads to reduced confidence among neuroradiologists in computerised tools.

Table 3.3 summarises the most common measures used to evaluate the MS lesion segmentation algorithms. We have distinguished between two main groups: those that evaluate a "hard" result (i.e. each voxel is assigned to only one tissue type) and those that provide a probabilistic result (i.e. each voxel has a membership value for belonging to the different tissue types). However, one should notice that all these measures are highly related. For instance, the probabilistic similarity index (PSI) is closely related to the Dice similarity coefficient (DSC).

Other measures used that are not related to the ones described in our table are those based on distance measures (in voxels or in millimetres). The aim of these measures is to

Table 3.3: Typical measures used when evaluating MS lesion detection.

	Name	Computation
Hard	Accuracy Percentage agreement	$\frac{\|TN\|+\|TP\|}{\|TN\|+\|TP\|+\|FP\|+\|FN\|}$
	Jaccard similarity index [47]	$\frac{\|TP\|}{\|TP\|+\|FN\|+\|FP\|}$
	Dice similarity coefficient (DSC) [10]	$\frac{2\times\|TP\|}{2\times\|TP\|+\|FP\|+\|FN\|}$
	Error rate	$\frac{\|FP\|+\|FN\|}{\|FP\|+\|FN\|+\|TP\|+\|TN\|}$
	Sensitivity Overlap Fraction Percentage of Correct Estimation True Positive Fraction/Rate	$\frac{\|TP\|}{\|TP\|+\|FN\|}$
	Specificity True Negative Fraction/Rate	$\frac{\|TN\|}{\|TN\|+\|FP\|}$
	False Positive Fraction/Rate	1 - Specificity
	Under estimation fraction (UEF) False Negative Volume Fraction	$\frac{\|FN\|}{\|TN\|+\|FN\|}$
	Over estimation fraction (OEF) Extra Fraction	$\frac{\|FP\|}{\|TN\|+\|FN\|}$
	Overlap objects fraction	$\frac{N_{obj}(TP)}{N_{obj}(Ref)}$
Soft	Probabilistic similarity index	$\frac{2\times\sum P_{x,gs=1}}{\sum 1_{x,gs=1}+\sum P_x}$
	Probabilistic overlap fraction	$\frac{\sum P_{x,gs=1}}{\sum 1_{x,gs=1}}$
	Probabilistic extra fraction	$\frac{\sum P_{x,gs=0}}{\sum 1_{x,gs=1}}$

evaluate how far the boundaries of an obtained lesion segmentation are from those of the real one. Although in general, these measures are not common in most of the analysed works, they were used in the MS Lesion Segmentation Challenge (2008) [75].

In conclusion, we would emphasise that, although DSC has become something of a standard measure for evaluating MS lesion detection methods, none of the measures are perfect for this purpose. In fact, as stated by Cardenes et al. [25], various different measures (i.e. measures based on intensity values, distances and connectivity) should be combined to obtain a more objective and reliable assessment.

3.3.3 Analysis of the results

In this section, we provide a qualitative comparison of the results obtained by the approaches we have analysed. Tables 3.5 and 3.4 summarise the synthetic and real data, the

evaluation measures and results obtained. As already mentioned, a quantitative evaluation among these approaches is a difficult task due to the variability in the data sets and in the measures used.

Tables 3.5 and 3.4 clearly shows that the most commonly-used evaluation measures are the DSC and sensitivity (also known as the overlap fraction). The results obtained using real data range from 0.47 to 0.808 for the Dice coefficient. The highest Dice coefficients reported involving a large amount of data were obtained by Sajja et al. [65], who used 23 volumes from which they obtained a mean Dice coefficient of 0.78 and de Boer et al. [28], who used 209 volumes from the Rotterdam Study obtaining a mean DSC of 0.72. The work presented by Van Leemput et al. [53] provides an example of how the variability of the ground truth affects the results obtained. Using the same automatically segmented data and comparing them with two different expert annotations, the DSC varies more than 10% (from 0.47 to 0.58). These results illustrate the usefulness of algorithms that combine the ground truth.

Several works also use sensitivity and specificity as evaluation measures (see table 3.4). Observe that the specificity reported by the different methods is always close to 1. This is because this measure evaluates the ratio between the number of voxels correctly classified as healthy divided by the number of voxels automatically classified as healthy. Therefore, considering that lesions are small spots within the whole volume, the specificity value always tends to be close to 1.

The last four methods shown in table 3.5 used the synthetic BrainWeb phantom (a simulation model) to provide quantitative results. As already mentioned, the BrainWeb simulation package helps to compare the performance of different approaches more effectively. For example, each of these four methods included simulations with 3% added Gaussian noise and a zero bias field. The results show that the approach of García-Lorenzo et al. [38] clearly outperforms the others, obtaining a DSC of 0.87. The methods of Freifeld et al. [37] and the two by Shiee et al. [69, 70] obtained DSC values of 0.77, 0.72 and 0.81 respectively under the same conditions. Furthermore, García-Lorenzo et al. [38] also report the difference between using synthetic data and using real MRI data when evaluating algorithms. Notice that when evaluating the same approach but using seven real volumes from the McConnel Brain Imaging Centre, the DSC drastically dropped to 0.55.

Finally, we want to mention in this section the quantitative results published during the 2008 MS Challenge [1] [75]. In particular, we have summarised the reported results

[1] All this information can be found in: http://grand-challenge2008.bigr.nl/and

Table 3.4: Summary of the real results presented in the analysed articles. For the real data we show number of slices (s) or volumes (v) and the origin of the database. The two results for Van Leemput et al. are obtained by using as a ground-truth the segmentation of two different experts, while Wu et al. distinguish the results between T2 lesions (T2) and black holes (BH). DSC stands for Dice Similarity Coefficient, and the $m \sim n$ means that there where between m and n slices.

Article	Real (Database)	Measures	Results
Kamber (1995) [48]	12x56s (Montreal Neurological Inst.)	Error rate	0.02-0.04
Goldberg (1998) [41]	14x10s (Sheba Medical Center)	Sensitivity	0.87
		Specificity	0.96
Bedell (1998) [12]	13v (Univ. of Texas Medical School at Houston)	Qualitative	-
Bourdraa (2000) [16]	10x22s (Hôpital d'Antiquaille)	Sensitivity	0.65
Alfano (2000) [2]	84x16s (Univ. Federico II)	Sensitivity	0.81
Van Leemput (2001) [53]	50v (BIOMORPH project)	DSC	0.47 & 0.58
Zijdenbos (2002) [87]	29v (Montreal Neurological Inst.)	DSC	0.60
He (2002) [44]	5v (Univ. of Texas Medical School at Houston)	Kappa	0.9
Anbeek (2004) [4]	20x38s (Univ. Medical Center Utrecht)	DSC	0.80
		Sensitivity	0.79
		OEF	0.19
Anbeek (2005) [3]	10x5s (Univ. Medical Center Utrecht)	DSC	0.808
		Sensitivity	0.815
		Specificity	0.999
Datta (2006) [27]	14v (Sanjay Gandhi Post-Graduate	DSC	0.73 ± 0.11
	Inst. of Medical Sciences)	Sensitivity	0.72 ± 0.13
		OEF	0.27 ± 0.21
		UEF	0.28 ± 0.13
Sajja (2006) [65]	23v (Sanjay Gandhi Post-Graduate	DSC	0.78 ± 0.12
	Inst. of Medical Sciences)	Sensitivity	0.88 ± 0.13
		OEF	0.38 ± 0.27
		UEF	0.11 ± 0.13
Wu (2006) [86]	6x2v (Slotervaart Hospital)	Sensitivity	0.70 (T2)
			0.62 (BH)
		Specificity	0.98 (T2)
			0.99 (BH)
Datta (2007) [26]	22v (Sanjay Gandhi Post-Graduate	DSC	0.76 ± 0.18
	Inst. of Medical Sciences)	Sensitivity	0.74 ± 0.22
		OEF	0.25 ± 0.62
		UEF	0.26 ± 0.22
Khayati (2008) [50]	20x12$\sim$20s (Koorosh Diagnostics	DSC	0.7504
	and Medical Imaging Center)	Sensitivity	0.7402
		OEF	0.2303
Garcia-Lorenzo (2008) [38]	7v (McConnel Brain Imaging Centre)	DSC	0.55 ± 0.05
Garcia-Lorenzo (2008) [40]	3v (MR and Image Analysis Research Centre)	DSC	0.56
Tomas (2009) [78]	9v (Boston Children's Hospital)	Sensitivity	0.65 ± 0.29
		FPR	0.71 ± 0.21
Subbanna (2009) [76]	10v (Montreal Neurological Inst.)	DSC	0.71
		FPR	0.00
		FNR	0.10
Akselrod-Ballin (2009) [1]	25+16vx24s (Scientific Institute Ospedale San Raffaele)	DSC	0.53 ± 0.1
		Sensitivity	0.55 ± 0.13
		Specificity	0.98 ± 0.01
		Accuracy	0.97 ± 0.01
de Boer (2009) [28]	209v (Erasmus University Rotterdam)	DSC	0.72
Shiee (2010) [70]	10v (National Multiple Sclerosis Society)	DSC	0.63

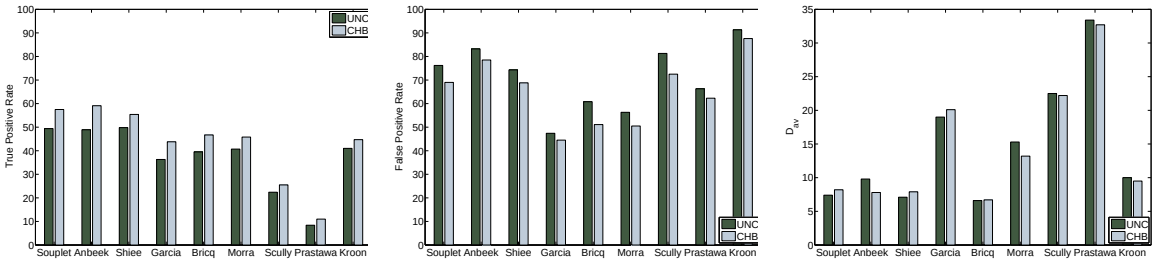
Table 3.5: Summary of the synthetic results presented in the analysed articles. For the Brainweb database we include the noise and bias field of the tested volume (n$x$bf$y$ means noise x and bias field y). The two results for Van Leemput et al. are obtained by using as a ground-truth the segmentation of two different experts, while Wu et al. distinguish the results between T2 lesions (T2) and black holes (BH). DSC stands for Dice Similarity Coefficient, and the $m \sim n$ means that there where between m and n slices.

Article	Synthetic	Measures	Results
Freifeld (2007) [37]	Brainweb (n3bf0)	DSC	0.77
	Brainweb (n9bf0)		0.73
Garcia-Lorenzo (2008) [38]	Brainweb (n3bf0)	DSC	0.87
	Brainweb (n3bf20)		0.85
	Brainweb (n3bf40)		0.63
Shiee (2008) [69]	Brainweb (n3bf0)	DSC	0.720
	Brainweb (n9bf0)		0.591
Shiee (2010) [70]	Brainweb (n3bf0)	DSC	0.812

in figure 3.7, which consists of three plots showing the results obtained when using three different evaluation measures: (a) the true positive rate (per lesion), (b) the false positive rate (per lesion) and (c) the average symmetric surface distance, which measures how far away the correctly segmented lesions are from the ground truth. The results in the plots have been ordered according to the final ranking in the on-site testing competition (from left to right). Notice that these plots illustrate slight differences in results when using the ground truth from the UNC experts or from the CHB centre. Slightly better performances were obtained when using the annotations from the CHB expert during training. As suggested by Styner et al. [75], this may be due to the fact that the UNC ground truth was obtained from two experts, while the CHB ground truth was obtained from only one. In fact, the training and testing processes for the UNC can be performed using ground truth from different experts while this is not possible at the CHB centre.

In general, figure 3.7(a) shows us that there is room for improvement regarding the true positive rates. The best performances were obtained by Souplet et al. [72], Anbeek [5] and Shiee et al. [68], each of whom had true positive rates of around 60%. There is also room for improvement regarding the false positive rates (see figure 3.7(b)), the best results being obtained by Garcia-Lorenzo et al. [39], Bricq et al. [18] and Morra et al. [58], each with false positive rates of around 50%. That is to say, from each pair of marked lesions, only one was correctly placed. However, as stated in [58], this problem can be successfully overcome by applying an automated post-processing step for false positive reduction. Finally, figure 3.7(c) shows the average symmetric surface distance. Using this measure based on the distance from the ground truth to the lesion surface contour, the

http://www.midasjournal.org/browse/publication/638



(a) (b) (c)

Figure 3.7: Results extracted from the works presented in the 2008 MS Challenge. Each plot details the results when using ground truth provided by either the UNC or the CHB experts. (a) shows the true positive rate per lesion, (b) the false positive rate per lesion and (c) the average distance between obtained and ground truth segmentations. The results are ordered according to the final ranking in the on-site testing competition.

best results were obtained by Souplet et al. [72], Anbeek et al. [5], Shiee et al. [68], Bricq et al. [18] and Kroon et al. [51], with results of between 5 and 10 mm. Considering that the nominal diameter of an MS lesion is about 7mm [84], the results obtained, although promising, are still far from the requirements of a perfect automated volumetric tool.

In addition to the results obtained from the manual segmentations, the MS Challenge organisers computed a composite segmentation using the well-known STAPLE algorithm [85]. Specifically, the input for STAPLE included all the manual segmentations, as well as the segmentations provided by the workshop participants. Hence, it represented a composite of two human experts and nine automated segmentation methods. With respect to this evaluation experiment, the best sensitivities (which were from the works of Anbeek et al., Morra et al. and Kroon et al.,) were around 0.5 while the specificities were close to 1, showing the ability of these algorithms to identify lesions correctly. These experiments using the STAPLE segmentation illustrate the improvement in reducing the false positive rate. In summary, the quantitative evaluation performed in the MS Segmentation Challenge 2008 [75] has revealed both the challenges facing the participants as well as the need to develop new approaches to MS lesion segmentation.

3.4 Discussion

As seen in previous sections, the most widely-used feature in all segmentation methods is voxel intensity, which is commonly employed with a multi-channel approach. In addition,

features based on modelling the voxel neighbourhood are also used in some approaches to introduce (local) spatial information to the algorithms. Regarding the modalities, we have seen that T1-w images are widely used for the tissue segmentation and also for the black holes and to enhance lesion segmentation. T2-w and PD-w images are typically used for detecting MS lesions. However, the major drawback of these images is the similarity in the intensities of lesions and CSF. Due to this similarity, the discrimination between ventricles and lesions may be difficult, especially when they are connected. Some approaches perform another segmentation step to solve this problem. FLAIR images also provide good discrimination between lesions and healthy tissue but, as some authors have pointed out, they have problems when dealing with sub-cortical structures due to image artifacts.

Most of the lesion segmentation approaches we have reviewed are based on iterative algorithms. Among them, the commonest technique is the EM, which obtains the model parameters in an efficient way. This technique also allows different models to be used for different tissues, which is indeed very useful since WM and GM can be assumed to follow Gaussian distributions while CSF does not follow any known distribution [65]. While iterative approaches are inherently unsupervised approaches (even though they may use a priori information), supervised approaches have also been proposed for the purpose of MS lesion segmentation. Most methods rely on pattern recognition techniques, the commonest being the kNN, to detect voxels that are either outliers to the tissue models or are similar to the lesion model (derived from a training set).

Atlas registration is another way of tackling MS lesion segmentation in brain MRI images. These approaches are useful since the lesion intensities often overlap with the intensities of other structures in the brain. Hence, atlases might provide valuable contextual information to eliminate possible false positives. As a drawback, the registration process is a computationally intensive procedure. Moreover, the results obtained by these methods are affected by the physiological variability of each subject and may lead to erroneous results in the case of diseased brains. This is because atlases are based on normal brains and lesions may appear almost anywhere in the atlas, making it difficult to construct atlases from diseased brains (although this has been analysed in [51]).

Analysing the results, we noticed that the best performances in terms of DSC and sensitivity (either using their own data sets or the MS Challenge data) were obtained by Anbeek et al. [4, 3, 5], Sajja et al. [65], de Boer et al. [28], Souplet et al. [72] and Shiee et al. [68]. What all these approaches have in common is a supervised strategy. In particular,

the work of de Boer et al. [28], Shiee et al. [68] and Souplet et al. [72] are based on using an Atlas to perform the segmentation, while the rest [4, 3, 5, 65] used manual segmentations to train the classifiers. Finally, the best results reported in the MS Segmentation Challenge were from the supervised strategy presented by Souplet et al. [72].

Chapter 4

Proposal

Automatic MS lesion segmentation is a challenging task due to MRI artifacts and the overlapping tissue intensity distributions. Moreover, our analysis of the most recent techniques presented in chapter 3 has demonstrated an improvement of the results obtained when using a priori information to introduce spatial information. Atlases built from manual segmentations can provide this prior tissue information after a registration process. This new step, while increasing the accuracy after a preprocessing stage, also increases the whole framework complexity. In this chapter, we will present a pipeline proposal (including preprocessing, registration, tissue classification and lesion segmentation) based on the best techniques of the current state-of-the-art.

4.1 Introduction

In this chapter, we present our proposal to segment tissues and MS lesions using MRI. Based on the state-of-the-art analysis, we have implemented a pipeline approach. Consequently, our proposal comprises 4 different stages as illustrated in figure 4.1, each one focusing on solving a different problem.

The first step consists on applying different pre-processing steps to enhance the image quality. Afterwards, the probabilistic atlases of each tissue are registered to the new MRI volume. Finally, the segmentation is carried out in two major stages. Using the atlas information, tissues are classified first. Then this tissue information is integrated into the lesion segmentation stage.

We decided to implement two different strategies for the tissue stage based on the works

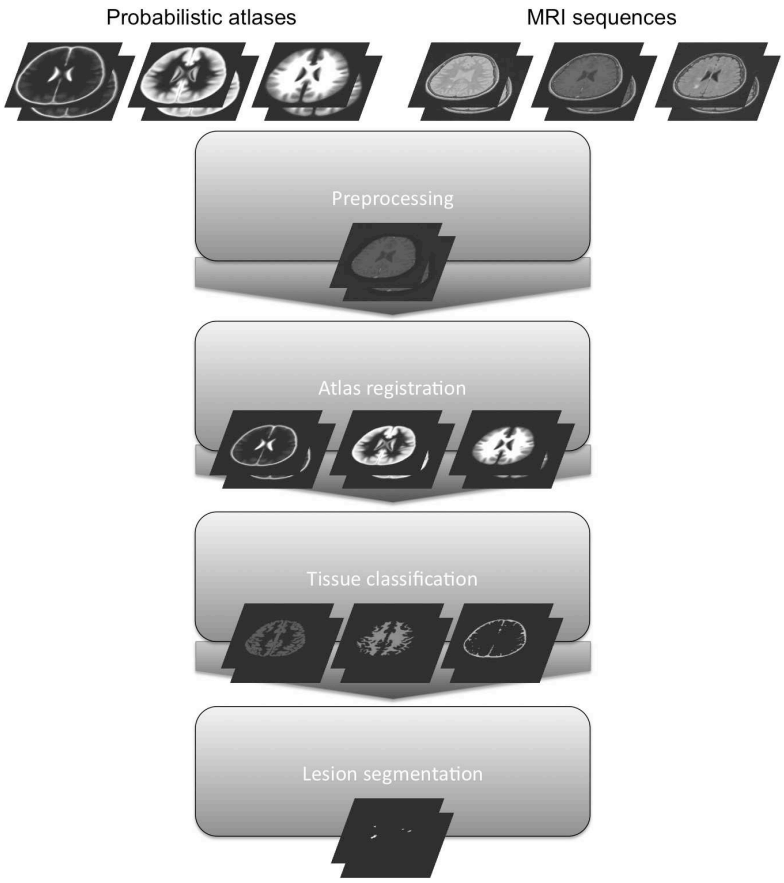
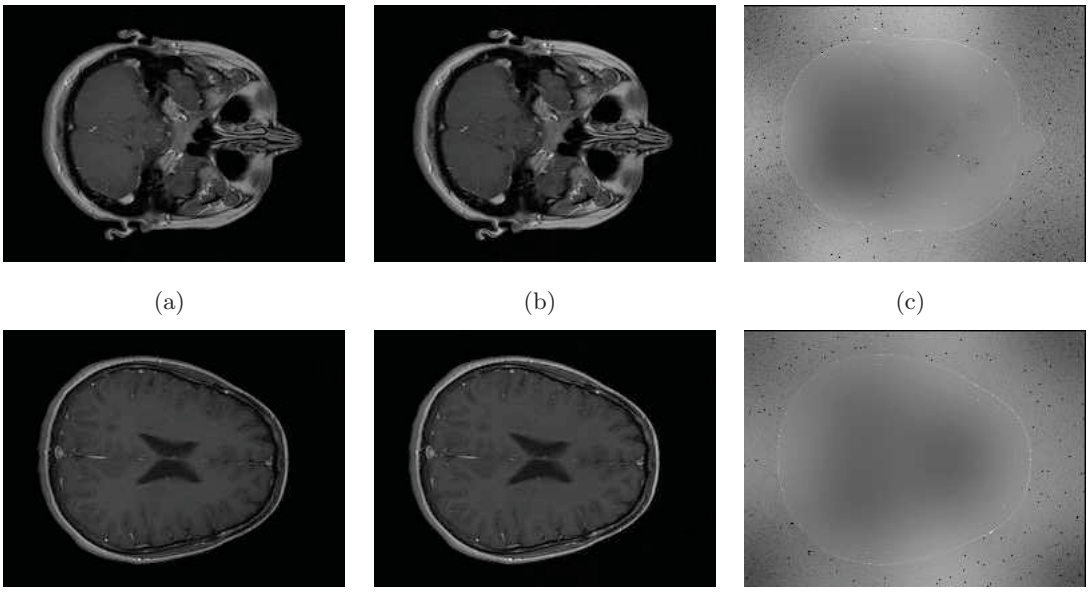


Figure 4.1: Scheme of our pipeline approach. Initially, different preprocessing steps are applied to enhance image quality. After that, the probabilistic atlases are registered to the new image. Those atlases are then used for the tissue segmentation. Finally, lesions are segmented using this tissue information.

of Souplet et al. (2008) [73] and de Boer et al. (2009) [28]. However, for the lesion stage both papers follow the same idea with different steps, therefore we decided to combine them. As stated in chapter 3, those approaches obtained the best results with large data sets of real patients.

In the next sections, we will present each part of this pipeline and we will describe our approach for each step.



(a) (b) (c)

(d) (e) (f)

Figure 4.2: Bias correction results on two different images. a) and d) The original images without preprocessing, b) and e) the same images after applying the SPM tools and c) and f) the bias field estimated by the method.

4.2 Preprocessing

Two major preprocessing steps are required to reduce missclassification errors. Those steps are bias correction and skull stripping. In our approach, we decided to use publicly available tools to perform those two steps:

- **Bias correction:** Two different studies have recently revised different ways to overcome this pre-processing step problem [46, 83]. Both studies classify these methods into various groups: segmentation-based, filtering-based, surface fitting-based, histogram-based and other specific techniques. However, as pointed out by Hou [46], none of the methods has been shown to be superior to the others and exclusively applicable. Therefore, we decided to use the SPM [8] toolbox for Matlab. One of its tools implements bias correction via segmentation after reducing also image noise. Results obtained after applying this preprocessing step for two different images are presented in figure 4.2.

- **Skull stripping:** Two recent works have analysed and compared the state-of-the-art methods to extract brain regions from MRI images. The first study, by Boesen

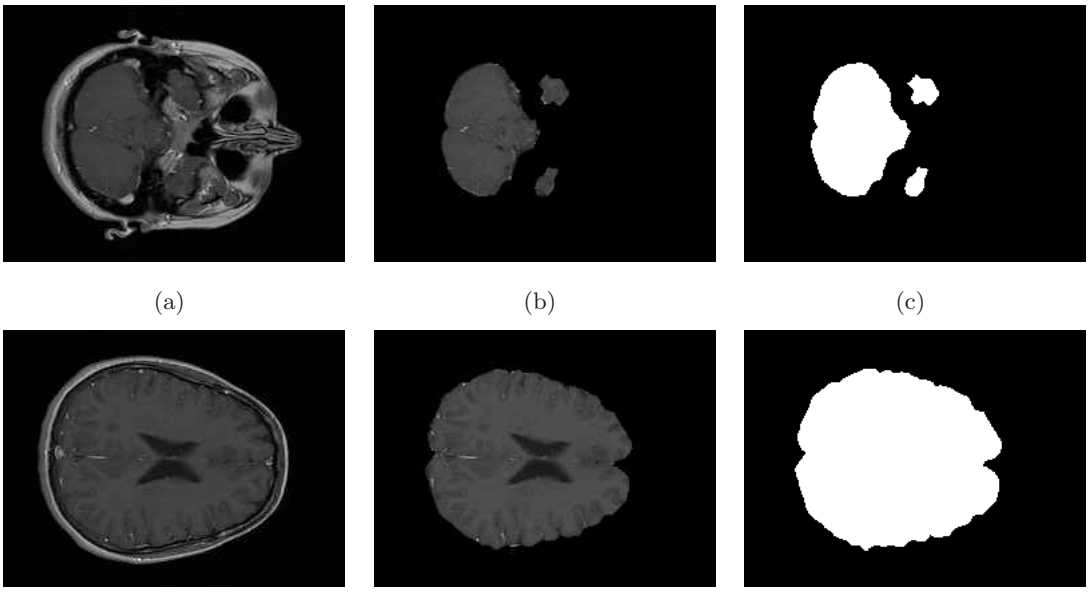


Figure 4.3: Skull stripping results on two different images. a) and d) Images after bias correction, b) and e) images after applying BET and c) and f) the brain mask.

et al. [15], compared four systems: statistical parametric mapping (SPM2) [8], the brain extraction tool (BET) [71], the brain surface extractor (BSE) [67] and their own Minneapolis consensus strip (McStrip) [62]. They validated these systems with three data sets of T1-w images. The second study, by Hartley et al. [43], compared only the accuracy of BET and BSE against 296 PD-w images. In both studies, manual segmentations were used as the "gold standard" Boesen et al. [15] concluded that the McStrip - which is a hybrid algorithm incorporating intensity thresholding, nonlinear warping and edge detection - consistently outperformed SPM2, BET and BSE.

Due to McStrip not being publicly available and its dependence to other methods, we decided to use the FSL tool called BET [71] to remove nonbrain tissue. This method fits a surface to brain tissues in an iterative way. Results obtained after applying this preprocessing step for two different images are presented in figure 4.3.

Those two preprocessing steps will also aid the registration process since most registration techniques rely on image intensities to evaluate the alignement between images at each step. Moreover, nonbrain tissues may guide the registration process outside brain tissues (such as the eyes) instead of focusing on adjusting GM and WM sulci, causing

nondesired deformations.

4.3 Registration

A registration step is needed in order to align the atlas with the new volume. Registration techniques can be divided into global and local. Global transforms provide an initial alignment between a reference volume, usually called the fixed volume, and the one being registered, usually called the floating or the moving volume. Those transforms can also assume deformations in the form of shearing or scaling but always maintaining some proportion on the volume as a whole. However, inner structures from diferent subjects can present severe deformations that can be solved using local transforms.

4.3.1 Global transformation

Rigid body transforms are the easiest way to align two different volumes, comprising solely a rotation on each axis and a translation to the origin. Since those transforms are in fact linear equations, they can be expressed by a single homogeneous matrix, where R_{ij} defines the rotation (based on 3 angles θ_x, θ_y and θ_z) and T_k defines the translation (in the 3 axes). Note that by only applying translations and rotations no deformation is defined so all distances between voxels are maintained.

$$M = \begin{pmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.1}$$

As a consequence of defining the rotation as a product of different rotations on each axis by θ_x, θ_y and θ_z orthogonality restrictions that might not be realistic are imposed to the global transform.

To relax those constraints, rigid transformations of 6 degrees of freedom can be extended as a 12 parameter transform, where no orthogonality restrictions are imposed. Those parameters implicitly include the same rotation and translation as well as a shearing and scaling deformation on each axis. Note, however, that those affine transforms affect each brain voxel in the same way, hence maintaining the transform global.

4.3.2 Local transformation

Rigid registration is usually enough when dealing with intrasubject medical applications, such as temporal studies of the same subject. However, when dealing with intersubject applications such as atlas matching nonrigid algorithms can explain local anatomical variations between the template and the subject brain that global methods fail to reproduce (i.e. GM sulci).

In general, volume partitioning is performed to account for these local deformations. For instance, the moving volume can be decomposed on smaller sub-volumes and those sub-volumes can then be rigidly registered to the fixed volume using the homogeneous matrix of equation 4.1. This method [45, 6] is hierarchical since first the whole volume is aligned using rigid transforms and then each divided sub-volume is again registered using rigid transforms.

Other partitioning approaches define a uniform grid on the whole volume and then apply to each of those grid vertices some non-linear transform that can no longer be defined as an homogeneous matrix. Depending on the accuracy needed and efficiency requirements in both time and space, grid vertices can be defined as the volume voxels. Common non-linear transforms based on mathematic transformations are cosine based functions [7], b-spline curves [64, 59] or level set partial differential equations [80]. Other functions defining the displacement fields have been proposed based on the thermodynamics concept of demons [77, 82] or optical flow models [81, 60].

Choosing the right registration technique is crucial on any atlas-based method. In fact, misalignments may change the final labelling and therefore cause misclassification errors. This misalignment is usually produced by inter-subject variability which can not be expressed in the form of global transforms. Consequently, local registration methods defining non-linear deformations have gained special interest in the last few years.

In our approach, we have decided to implement with ITK[1] one of the most common methods based on a multi-resolution approach [64] with a starting pre-registration using affine transformations to globally align the atlas template and the patient's scan, followed by a b-spline transformation using a grid of 32 control points per dimension. This value has been empirically set up in our experimental results as a tradeoff between performance and computational cost.

[1] This library is publicly avilable in http://www.itk.org

4.4 Atlas-based segmentation

As stated on previous chapters, automatic segmentation of brain MRI images is a challenging task due to the inherent inhomogeneities of MRI data and the fact that different structures can share the same grey-level intensities. Hence *a priori* anatomical information is essential to perform the segmentation task. This prior information may be provided in different ways. For instance, as a set of rules based on tissue properties, as a set of regions of interest, or as a set of manual expert delineations. Actually, those delineations can be considered as an anatomical brain **atlas** and can be used as a reference to guide the segmentation algorithm after registration to the new case space.

Thus, atlas-based segmentation is defined as the process of delimiting the brain tissues and/or structures using this a priori anatomical model. In this section, we diferentiate between four different strategies (label propagation, multi-atlas fusion, feature definition and sample creation) to profit from this a priori information, briefly described with advantages and drawbacks in table 4.1.

The easiest and fastest way to assign a label to each voxel after registration is to propagate atlas labels since it only relies on this single registration. However, missclassifiation errors may arise in cases where a large discrepancy exists with the atlas. Moreover, creating a single generic model robust enough against new subjects and specially when dealing with abnormal types in random locations is not an easy task.

This technique is extended by using multiple atlases to account for variability. Furthermore, outliers are minimised since those voxels with low agreement between different manual segmentations can be discarded. Due to those strengths, the best results on structure segmentation were obtained using multi-atlas fusion. However, other issues arise like atlas selection (to discard those templates that differ from the subject) and their combination, besides, a time complexity increase due to a higher number of registrations. Furthermore, those methods can not be directly applied to abnormal types since only atlas labels are propagated.

If statistical atlases are used instead, voxel probabilities can be integrated as part of a statistical framework and also as input features. Using this strategy, atlas values do not represent final segmentation but define a probability (membership) for each voxel. However, those probabilities might not completely represent the underlying segmentation, hence atlas values must be weighted to avoid bias. As a consequence, complex statistical models have to be estimated, affecting the final time complexity. In contrast, only a single

registration is needed and these complex models can also be used to segment abnormal lesions, considering voxels far away from the model as outliers. In fact, the best results on healthy and abnormal tissue are obtained using atlases as features.

Since atlases might not match correctly to the subject, they can also be used to estimate a robust subset of volume voxels as samples for classification or distribution estimation. By registering a single statistical atlas, variability is still accounted for and outliers can be discarded during sample estimation. However, how those robust samples are estimated using also intensity values supposes a challenge by itself and increases time complexity. On the other hand, using robust estimation and outlier strategies, abnormal tissue types can also be sampled with good results on the final segmentation.

Looking at table 4.1 it is clear that the best strategies for abnormal tissue types (MS lesions) are those based on sample creation and feature definition. Based on that, we have chosen and implemented two of the best techniques of the state of the art following those strategies: a kNN approach from de Boer et al. [28], which has the best results on real data using a large data set (see table 3.4); and an EM approach from Souplet et al. [72], which ranked the best on MS challenge 2008 (see figure 3.7). Both strategies are actually an extension of tissue segmentation strategies with a final step consisting on segmenting lesions using FLAIR images.

4.4.1 Tissue classification with kNN

Based on the approach of de Boer et al. [28] CSF, GM and WM are segmented using an automatically trained kNN classifier which is an extension of the work by Cocosco et al. [21]. Training samples for the kNN classifier are obtained from the subject itself by atlas-based registration [64].

After registering the atlases to the new image using first an affine transformation and then a dense b-spline deformation, those atlases are thresholded in order to get candidate training samples with a predefined probability to belong to a specific label. A threshold of 0.7 is chosen in order to include both sulcal (external) and ventricular (internal) CSF, as shown in picture 4.4.

For all three classes, 7500 candidate training samples are randomly taken from the spatial locations masked by the thresholded atlases. The features of the samples consist of the intensity values of the PD-w and T1-w images at the sample locations. A pruning step is applied to the initial set of samples to remove samples with incorrect labels.

Table 4.1: Overview of the different atlas-based strategies with advantages and drawbacks.

	Description	Strengths	Weaknesses
Label propagation	Atlas labels are directly applied to the subject voxels	Intuitive, fast and easy Good for ROI definition Depends only on registration A single registration is needed	Errors with cases deviating from the model Hard to create a generic atlas Can not be directly applied to abnormal types
Multi-atlas fusion	Multiple labels are combined in the subject voxels	Accounts for subject variability Minimises outliers Good results with structures	Selection of the best candidates is needed Combination after registration is needed Multiple registration are needed Can not be directly applied to abnormal types
Feature definition	Atlas probabilities or labels are used as input to a segmentation framework	Accounts for subject variability Compatible with statistical theory Outlier detection can be used for abnormal types A single registration is needed Good results with tissues and abnormal types	Atlas weighting is needed to avoid bias Complex statistical model estimation is needed
Sample creation	Samples from the subject are selected	Accounts for subject variability Minimises outliers Outlier detection can be used for abnormal types A single registration is needed Good results with abnormal types	Robust sample estimation is a challenge Sample estimation requires extra time

(a) (b) (c)

Figure 4.4: Thresholded atlases at different values. CSF appears red, GM appears orange and WM appears yellow. a) Binary mask at t = 0.7, b) binary mask at t = 0.8 and c) binary mask at t = 0.9.

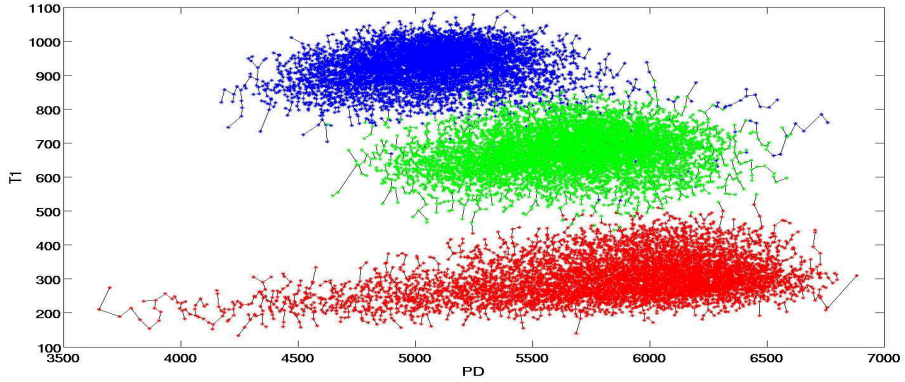
First, a minimal spanning tree (MST) of the samples in feature space is created as shown in figure 4.5. In an iterative process, the pruning algorithm removes connections whose length exceeds a threshold value equal to a constant multiplied with the average length of the other connections of a sample. At every iteration the threshold value decreases. This process is continued until every tissue class has a unique main cluster in feature space. A main cluster is defined as the cluster containing more samples of a certain class than the other clusters. A cluster is a unique main cluster when it is the main cluster for a single class. The final step removes all samples that are not connected or that are not in their main cluster.

Finally, a kNN classifier based on a *fast nearest neighbor lookup library*[2] performs the final classification based on the pruned sample set. It was shown [74, 30] that if k satisfies both equations 4.2 and 4.3 the kNN classifier will have the minimum possible error probability of a generic classifier (achieved when the classifier knows the true data distributions).

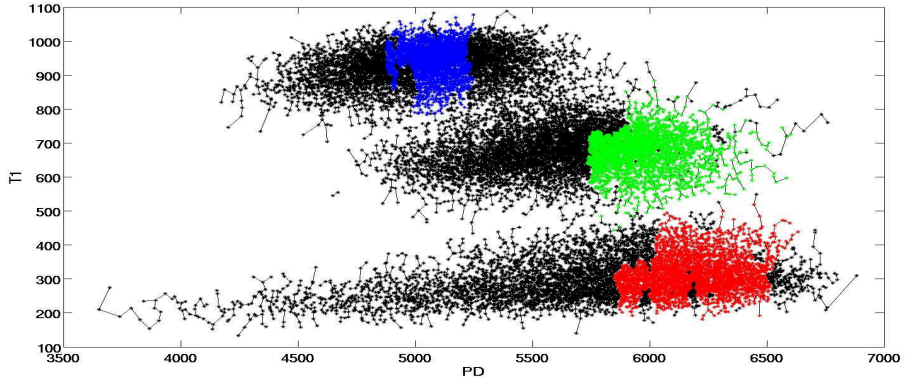
$$\lim_{n\to\infty} k = \infty \tag{4.2}$$

$$\lim_{n\to\infty} \frac{k}{n} = 0 \tag{4.3}$$

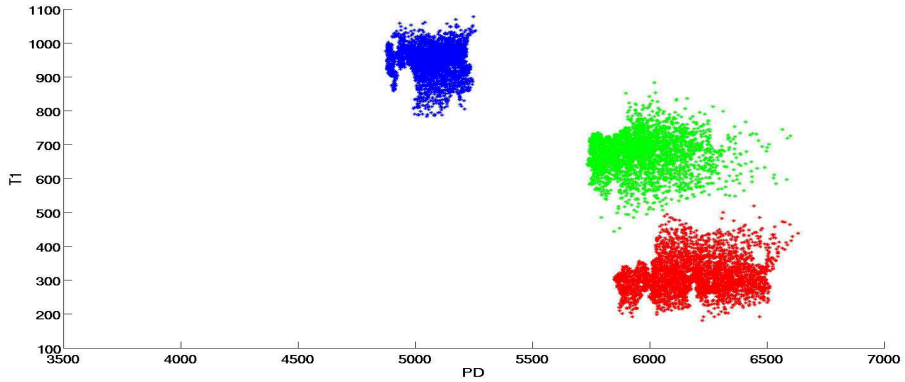
[2]This library can be downloaded from http://www.cs.umd.edu/mount/ANN/



(a)



(b)



(c)

Figure 4.5: Prunning stage example. a) MST of the initial set of samples for each tissue: GM (green), WM (blue) and CSF (red), b) MST after prunning and c) final set of samples for each tissue: GM (green), WM (blue), CSF (red) and outliers (black).

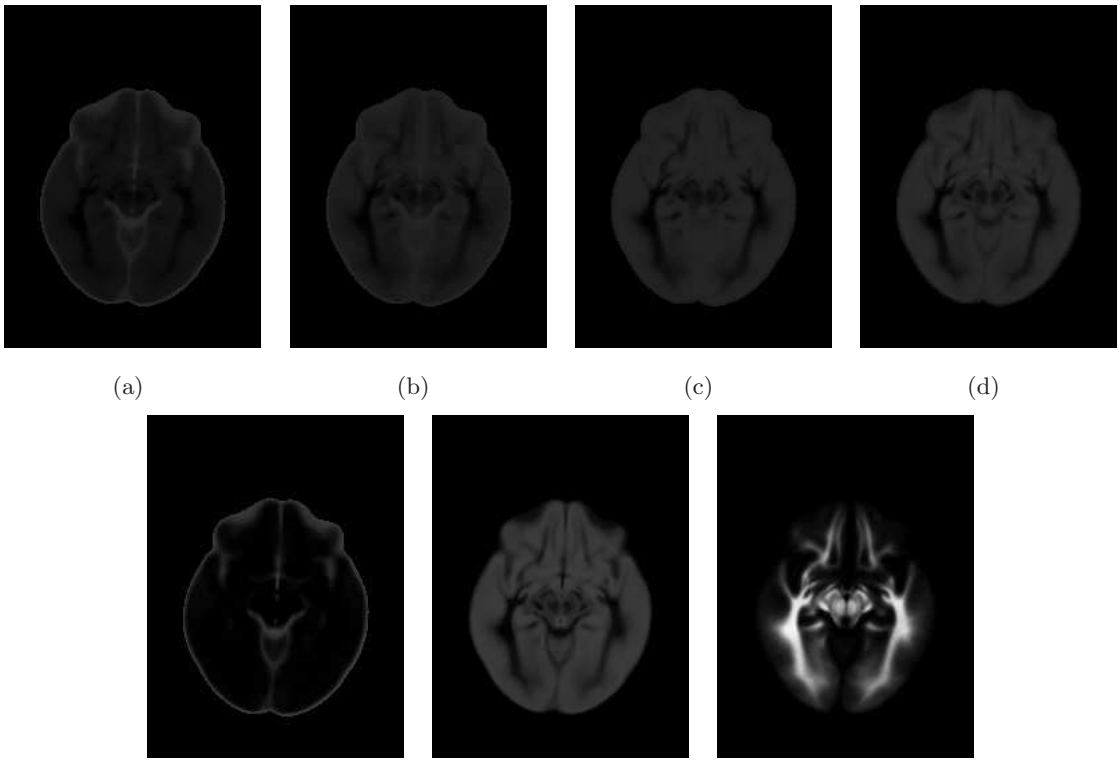


Figure 4.6: Atlases example including partial volumes. a), b), c) and d) 4 partial volume atlases with different proportions of CSF and GM, e) pure CSF atlas, f) pure GM atlas and g) pure WM atlas.

For example, $k = \sqrt{n}$ satisfies both conditions. Therefore, we decided to choose $k = 45$ and $n = 7500$ ($k \approx n^{0.43}$) similar to Cocosco et al. [21] and de Boer et al. [28].

4.4.2 Tissue classification with EM

Based on the approach of Souplet et al. [72], the algorithm presented in the work of Dugas-Phocion et al. [33] is applied on the T1-w and T2-w images. This algorithm uses the principle of the EM algorithm to maximise the log-likelihood between the MRI data and a gaussian model of ten classes: WM, GM, CSF, six GM/CSF partial volume classes (with different proportions), and an outlier class (additional class that corresponds mainly to the vessels).

First, the probability of belonging to the different classes of each voxel is initialised thanks to the a priori information of the atlas. Atlas probabilities are taken directly

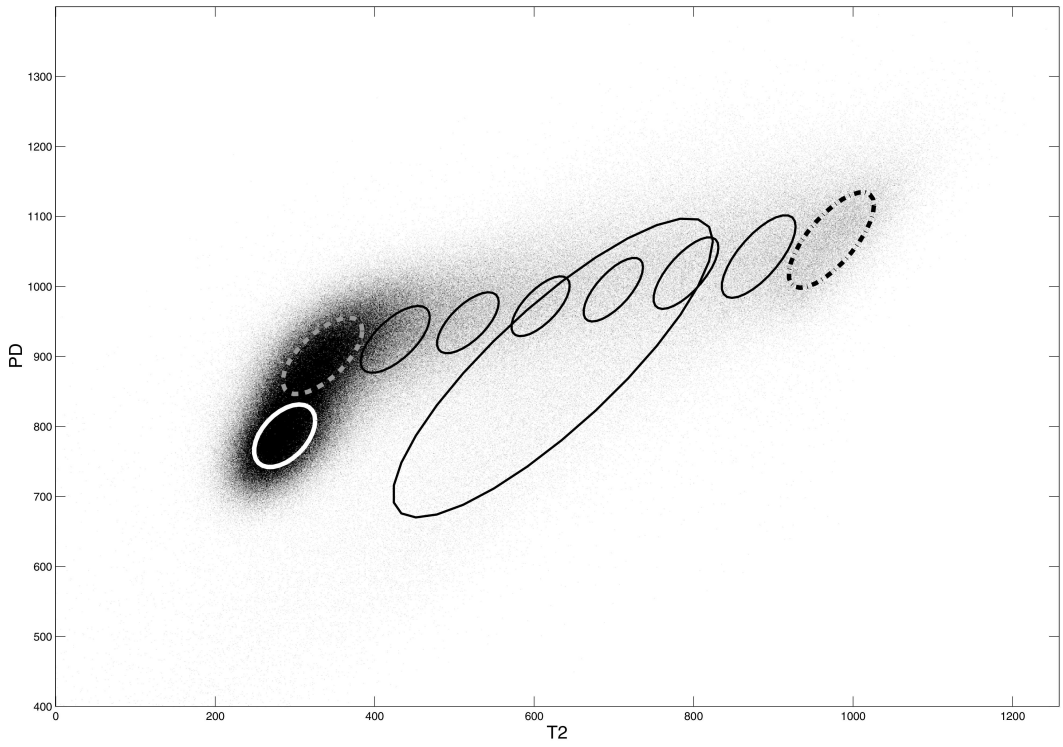


Figure 4.7: Example of the estimated Gaussian distributions represented by Mahalanobis ellipses. 10 distributions are estimated: GM (grey dashed), white matter (white), CSF (black dash-dotted), partial volumes (small black) and outliers corresponding to vessels (big black). Image extracted from the paper of Dugas-Phocion et al. [33].

from the atlases for pure classes (CSF, GM, WM), however partial volume classes have no priors. To fix this issue, new atlases with proportions of the CSF and GM atlases are created for the partial volume classes as shown in figure 4.6.

According to Dugas-Phocion et al. [33], the vessels class is located inside the CSF, therefore its atlas is defined also as a proportion of the CSF atlas. This proportion is estimated at each step. Finally, probabilities are normalised, since the sum of all the priors for each voxel should be 1.

After initialisation, two steps are iterated:

- In the maximisation step, the parameters (mean μ_k, covariance matrix S_k) of each class k are computed from the voxels intensities and their probabilities of belonging to the different classes. Partial volume classes parameters are obtained as a proportion

of the pure tissue parameters.

- In the expectation step, the probability of belonging to the different classes of each voxel is updated depending on the classes parameters, using the gaussian function and the new atlases values as prior probabilities.

Finally, when the algorithm converges, a bidimensional Gaussian distribution is estimated for each class including GM, WM, CSF, partial volumes and outliers as well as all the voxels probabilities related to those distributions. An example of those estimated distributions can be seen in figure 4.7.

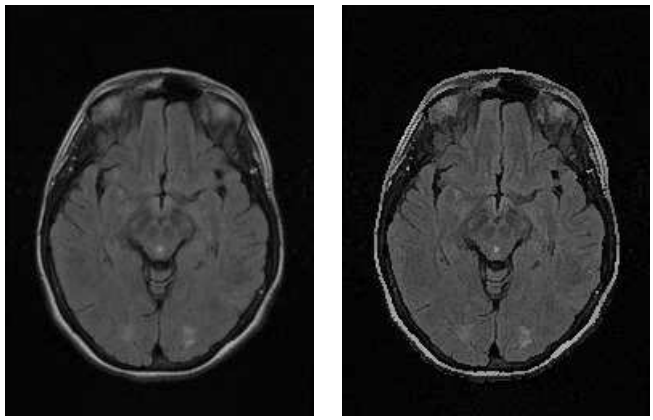
4.5 Lesion segmentation with tissue information

Based on a combination of the approaches by Souplet et al. (2008) [72] and de Boer et al. (2009) [28], the binary segmentations of each tissue of the brain are applied on the T2-FLAIR sequence to compute the properties (mean μ and standard deviation σ) of healthy compartments on this sequence. Both approaches segment lesion using the same technique with different pre- and post-processing steps (i.e. FLAIR image enhancement or FP reduction). As lesions are hyper-intense signals on the FLAIR sequence, a sensitive threshold T which gives a preliminary segmentation of the lesions can be estimated automatically from the properties of the GM class with equation 4.4.

$$T = \mu_{GM} + 2\sigma_{GM} \tag{4.4}$$

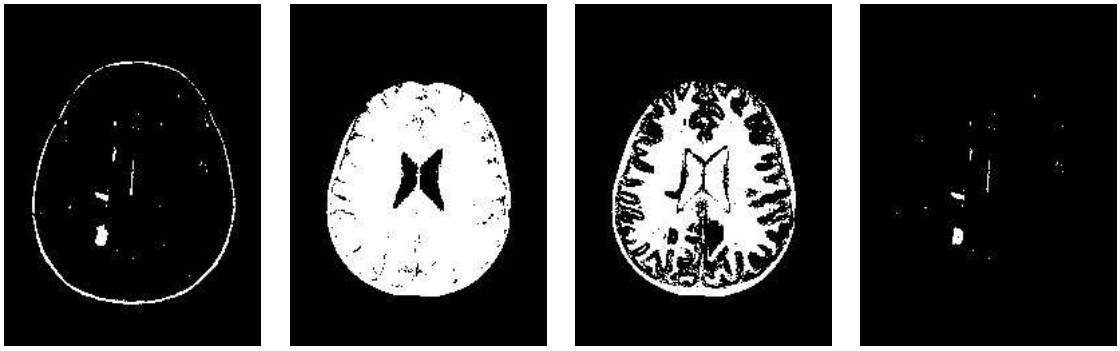
The application of this threshold on the T2-FLAIR sequence can help us to segment lesions (most of the lesion have at least a voxel with an intensity higher than the threshold). However, lesion voxel intensities are inhomogeneous and the "delineation" of the lesion is not simple even if a voxel of this lesion is known. For this reason, we enhance the contrast in the T2-FLAIR sequence before applying the threshold T by applying morphological greyscale operations (dilate and erode) as illustrated in figure 4.8.

This preliminary lesion mask may contain voxels belonging to WM or CSF. Moreover, this mask represents hyperintense spots that may be located outside the brain due to skull stripping errors. For this reason, we introduce two masks: a positive mask defining a region of interest where lesions should be located (mainly the WM mask from the tissue classification step with all holes filled) and a negative mask defining those regions where lesions can not appear (basically the WM and CSF mask from the tissue classification



(a) (b)

Figure 4.8: Contrast enhancement example: a) original FLAIR image and b) FLAIR image after contrast enhancement.



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

Figure 4.9: MS lesion refinement step. a) Initial lesion mask, b) positive mask, c) negative mask and d) final lesion segmentation.

step). Applying these masks outliers belonging to other tissues are removed from the lesion mask. Finally, voxels belonging to the final lesion mask are relabelled on the final segmentation as lesions. An example of this whole masking step is illustrated in figure 4.9.

Chapter 5

Results

MS lesion segmentation approaches are usually evaluated using different quantitative measures and both synthetic and real MRI volumes. In this section, we first present our datasets (probabilistic atlases, synthetic phantoms and real cases) followed by the results of our experiments. Finally, we compare and discuss the results presented by the different approaches, highlighting the most interesting aspects.

5.1 Datasets

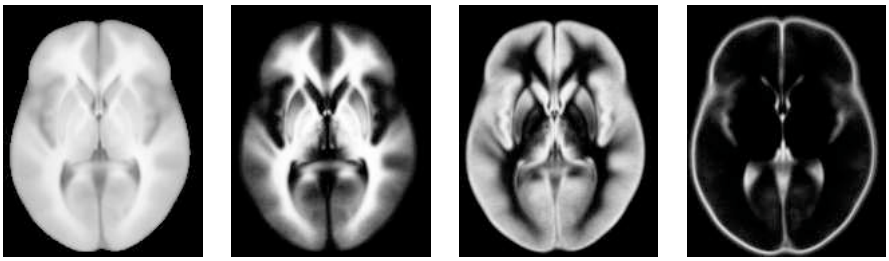
One of the main difficulties when comparing MS lesion segmentation approaches is the lack of a common database (with ground truth data) for training and testing the algorithms. Fortunately, this is becoming less of an issue with the appearance of new databases designed to meet this particular objective. In this section, we present a publicly available probabilistic atlas that we decided to use, as well as a public synthetic data set and our own database of real cases.

5.1.1 Atlas

A widely used atlas database is the one provided by the International Consortium for Brain Mapping (ICBM)[1]. This institution is focused on the ongoing development of a probabilistic reference system for the human brain. Nowadays, they have developed a tissue probabilistic atlas of $149 \times 188 \times 148$ voxels.

This atlas was build using 452 T1-w images of healthy brains and their manual segmen-

[1] http://www.loni.ucla.edu/ICBM/



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

Figure 5.1: Slice example of the ICBM452 atlas. From left to right: a) T1-w mean, b) WM probability map, c) GM probability map and d) CSF probability map.

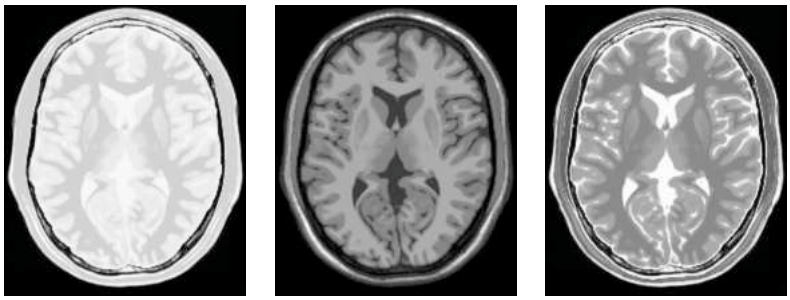
tations for each tissue (GM, WM and CSF). Those segmentations were then registered to a common space using only affine transformations (to prevent major deformations). Afterwards, they averaged all the T1-w images to create a reference template. Finally, probabilistic tissue atlases were built using the voxel frequency of being labelled as each tissue. Figure 5.1 shows the tissue probabilities for one central slice as well as the T1-w template.

We decided to use this atlas for two main reasons: 1) it is a publicly available resource widely used in this research topic and 2) this atlas is based on a large database of healthy brains. In order to profit from this variability, we first estimate the transformation parameters between the new T1-w image (used as the fixed volume) and the T1-w mean of the atlas (used as the moving volume) to finally apply those transformations to the tissue atlases that share space coordinates with the T1-w mean template.

5.1.2 Synthetic data

As mentioned in previous chapters, one of the main difficulties when comparing MS lesion segmentation approaches is the lack of a common database (with ground truth data) for training and testing the algorithms. Fortunately, new databases and synthetic phantoms are being developed in order to solve this issue.

Synthetic databases are useful as an initial framework for testing the various approaches to segmentation. BrainWeb [22] is becoming a standard synthetic database for evaluating both tissue classification and MS lesion segmentation algorithms. BrainWeb contains simulated T1-w, T2-w and PD-w brain MRI data based on two anatomical models: 1) a normal brain and 2) a brain containing MS lesions (see figure 5.2). Note that FLAIR



(a) (b) (c)

Figure 5.2: Brainweb slice of the lesioned phantom. From left to right: a) PD-w image, b) T1-w image and c) T2-w image.

simulations are not provided.

The MS model contains lesions segmented from real scans and adapted to the healthy model. This model with lesions is also becoming an standard for MS segmentation evaluation. Notice that the healthy phantom, based on an statistical tissue model of 305 MRI volumes, can also be used as a probabilistic atlas.

Moreover, this database provides different models according to parameters such as slice thicknesses, noise levels and levels of non-uniformity intensity. Default parameters are 3% noise, 20% bias field (based on the highest intensity values) and $1mm^3$ isotropic voxels, defining a volume of $181 \times 217 \times 181$ voxels. These data sets are available in three orthogonal views although the majority of algorithms use only the axial view.

5.1.3 Real data

As stated in chapter 1, we are collaborating with relevant hospitals and medical expert teams in the field of multiple sclerosis. Specifically:

- From the Hospital Vall d'Hebron (Siemens): Dr. Rovira, director of the "Unitat de Ressonància Magnètica-Centre Vall d'Hebron", (URMVH), has participated in several research projects funded by public and private institutions in the last few years. This group is part of and participates with the MAGNIMS [2], a European network of centers which share interest in the MS study through MRI.

- From the Clínica Girona (General Electrics): Dr. Barceló and Dr. Vilanova are the

[2](www.magnims.eu)

codirectors of the "Unidad de resonancia magnética" from the Clínica Girona and they are members of several national and international radiology societies.

- From the Hospital Josep Trueta (Philips): Dr. Ramió is the current coordinator of the "Unitat de Neuroimmunologia i Esclerosi Múltiple" as well as of Drs. Ana Quiles and Gemma Laguillo, who work for the radiology research group.

This collaboration brings us the opportunity to create a MS MRI database with images from different scanning machines (Siemens, General Electrics and Philips). From each hospital, we are gathering images from 15 patients under temporal studies. Those studies comprise an initial scan and its follow-up (after either 6 months or a year) and include the 4 main conventional MRI techniques described in chapter 2: PD-w, T1-w, T2-w and FLAIR images (see figure 2.2). The number of slice per volumes ranges between 44 and 46 according to an slice thickness of $3mm$, as shown in figure 5.3. Slice resolution ranges between 256×256 and 192×256 pixels. This difference is due to the use of vertical black lines to create a square image.

Right now, we have already collected 6 real volumes of different patients including the 4 sequences of the initial scan, and the control after a year. However, we still need the manual segmentations provided by the experts. These expert annotations (used as ground truth) are needed to perform a quantitative evaluation. As a consequence, we will present real results in a qualitative manner, comparing different aspects of the pipeline such as the transformation chosen or the tissue segmentation strategy used.

5.2 Synthetic results

Brainweb's database contributes to lesion segmentation evaluation providing a phantom with MS lesions. Using this phantom, segmentation methods can be validated before they are used with real cases. Moreover, this phantom can also be used to compare tissue segmentations. Therefore, we decided to test our pipeline approach with this data set to check the differences between the tissue segmentation strategies, and how those differences affect the final segmentation comparing our results with the ground truth using the DSC similarity index and the true positive fraction (TPF) introduced in chapter 3. The image data from the Brainweb was simulated with 3% noise and 0% bias field (n3bf0). As atlas priors we decided to use the fuzzy healthy phantom. Using this phantom we can skip the registration step avoiding also registration errors that may change the algorithms behaviour. As a consequence, we do not evaluate in this section how the chosen

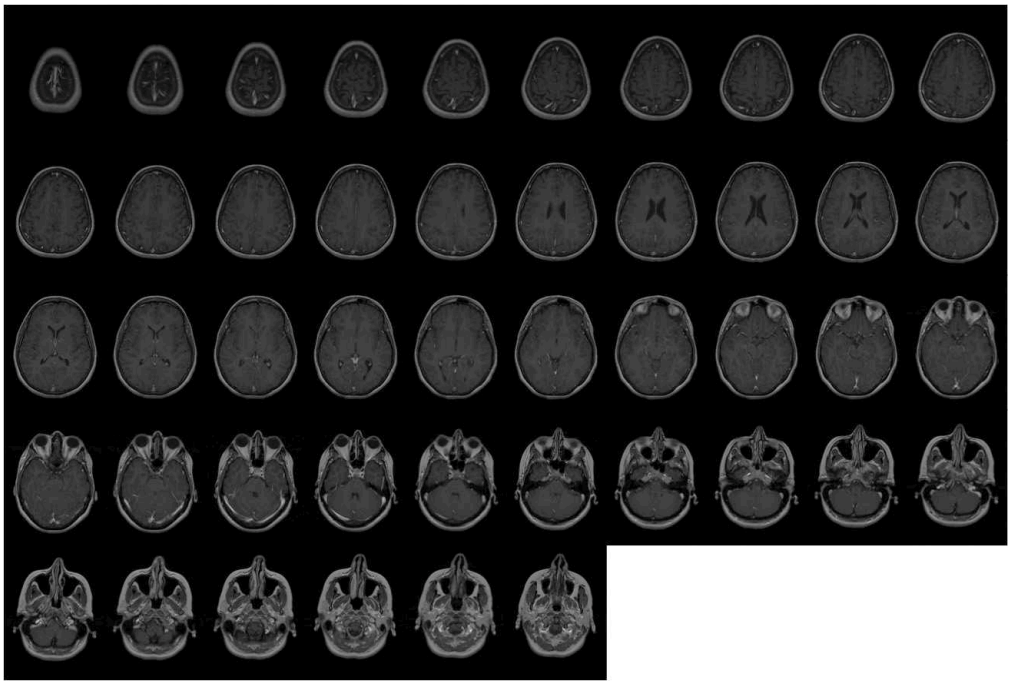


Figure 5.3: Example of all the 46 slices of a T1-w MRI volume.

transformation affects the final segmentation.

However, some issues need to be dealt with. As stated in chapter 4, our lesion segmentation is carried out using a FLAIR image. Since this sequence is not modelled by the Brainweb, we decided to use the T2-w image instead (that also keeps some contrast between GM and lesions) with CSF darkened to simulate the abscense of CSF on FLAIR images. Moreover, vessels and partial volume effects (PVE) are not modelled by this phantom. As a consequence, we tunned the parameters in order to adapt our approach to the images.

5.2.1 kNN vs EM

Synthetic results in terms of DSC and TPF are summarised in table 5.1. If we compare the DSC values obtained by both methods for the tissue segmentation, the EM-based strategy clearly outperforms the kNN one. Moreover, our approach based on the strategy of Souplet et. al. [72] also performs better than other state-of-the-art approaches presented in chapter 3 with DSC values over 90% for healthy tissues (see for instance the works of Shiee et al. [69, 70]).

Table 5.1: Synthetic results for lesion and tissues using the Brainweb (n3bf0). All results are obtained using the DSC and TPF evaluation measures. '-' stands for not available.

	WM	GM	CSF	WML
Freifeld (2007) [37]	-	-	-	0.77
García-Lorenzo (2008) [38]	-	-	-	0.87
Shiee (2008) [69]	0.92	0.90	0.90	0.72
Shiee (2010) [70]	0.93	0.91	0.90	0.81
SPM8 T1-w	0.88	0.90	0.65	-
SPM8 T2-w	0.86	0.84	0.70	-
SPM8 PD-w	0.84	0.79	0.68	-
kNN	0.88	0.79	0.75	0.27
EM	0.96	0.93	0.96	0.72

We also compared our method to the SPM software toolbox widely used by expert radiologists to perform the tissue segmentation. As mentioned briefly on chapter 4, SPM performs healthy tissue segmentation using atlas information and registration techniques while correcting the bias field simultaneously without taking lesions into account. Table 5.1 shows how our EM implementation also outperforms this software, while the kNN approach obtains similar values. It is important to remark that SPM does not segment lesions automatically therefore lesion voxels are misclassified decreasing its accuracy compared to our approach.

Regarding the lesion segmentation results from table 5.1, we can see that our EM approach also outperforms our kNN approach, due to a better prior tissue segmentation. Recall that our lesion segmentation approach relies on tissue segmentation to define a proper FLAIR threshold and proper positive and negative masks. Therefore, having good tissue masks and FLAIR images is crucial for obtaining good results. Nonetheless, our EM approach without a real FLAIR image obtains comparable values to other state-of-the-art approaches even though our results are fairly lower. In order to obtain those results, we applied a post-processing step based on morphological greyscale operations to the output of our pipeline to remove small lesions causing false positives. It is a common procedure to ensure a minnimum lesion size [72].

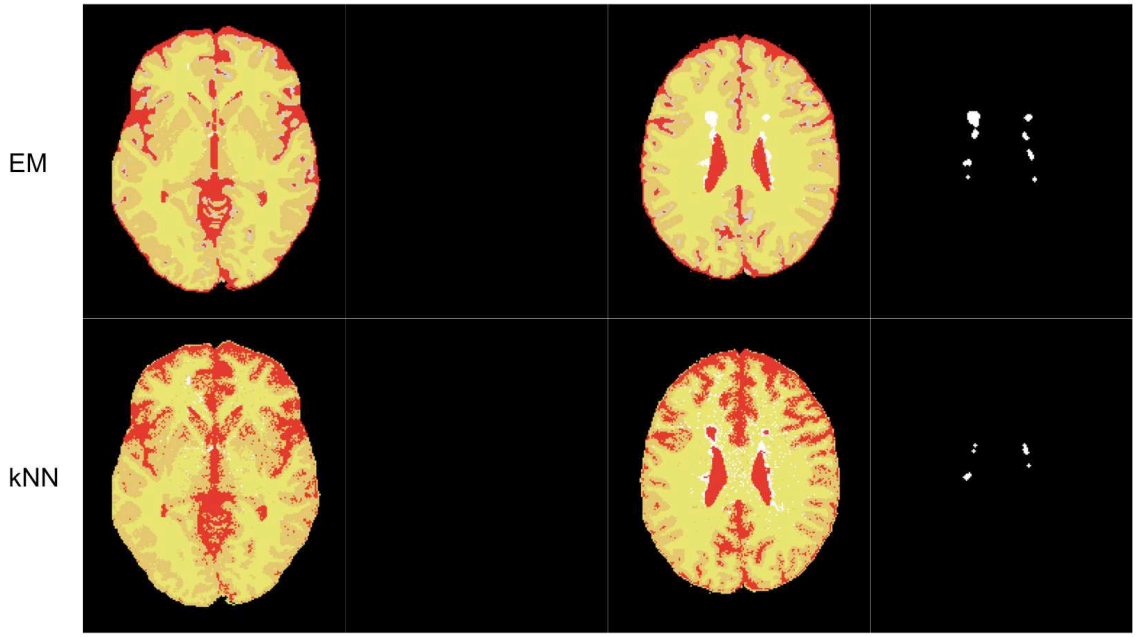


Figure 5.4: Brainweb segmentation examples. Segmentations for the EM strategy (first row) and the kNN strategy (second row) for GM (orange), WM (yellow), CSF (red) and lesions (white) for two different slices.

5.2.2 Discussion

Looking at figure 5.4 we can see the segmentation of MS lesions caused by both methods. Moreover we can see some differences between the proposed methods. In this section we will discuss those differencies and their causes while taking a deeper understanding on the mechanics of those two algorithms.

Our EM implementation assumes the intensities for each tissue are randomly taken from a Gaussian distribution. Besides, this implementation also assumes there are partial classes with a proportion from two different tissues (GM and CSF). In fact, Brainweb intensities follow a Gaussian distribution for each tissue. Recall that those images are synthetic simulations, therefore lacking real MRI artifacts. As a consequence, this strategy can accurately fit Gaussian distributions to each tissue obtaining clean tissue masks with few voxel missclassifications.

On the other hand, our kNN implementation does not assume any known distribution for the input data. Nonetheless, the training samples do define some intensity distribution depending on the parameter k. Therefore, choosing the right samples is crucial to guaran-

tee a good classification. Our implementation, selects random samples from a thresholded mask and then tries to estimate a robust subset of those samples. As a consequence, this subset may remove significative tissue samples considered as outliers. If we look at figure 5.4, we can see how some lesions are misclassified as CSF.

Summarising, the fact of using a synthetic phantom with clear Gaussian distributions influences positively in the EM segmentation but also decreases the kNN strategy accuracy when some samples are removed. However, kNN results should be similar to EM results when applied to real data where intensities do overlap and Gaussian distributions are not easily estimated. Moreover, the final lesion segmentation should improve using real FLAIR images where lesions appear clearly hyperintense compared to brain tissues.

5.3 Real results

Validation based only on synthetic images is not enough due to the difficulty of emulating real world scanning machines. For instance, PVE are not modelled in the Brainweb phantom and as a result, intensity values are clearly different among the three main tissues. Therefore, we decided to test our algorithms using real volumes. However, as stated before, we only provide a qualitative analysis of our experimental results. Notice that obtaining these manual segmentations is a time consuming task prone to errors.

5.3.1 Registration technique choice

In order to evaluate the segmentation step alone, we presented synthetic results with perfect alignment between the atlas and the image. However, when dealing with real cases, atlases have to be aligned using registration methods. Those methods, as stated in chapter 4, are time-consuming and depend on the transformation type used. Therefore, we evaluated the impact of this registration step with our real data sets when using only a global affine transformation or when using also a b-spline local transformation.

Comparing the strategies according to global affine transformations as shown in figure 5.5, we clearly see how the EM strategy outperforms the kNN one in terms of segmentation. Both methods are capable of detecting the lesions but lesions present holes using the strategy from de Boer et al. [28]. However, tissues are better outlined using the kNN approach since there are no PVE classes affecting pure tissue masks. Reducing the number of PVE classes and benefiting from spatial information would reduce this issue when using the EM approach.

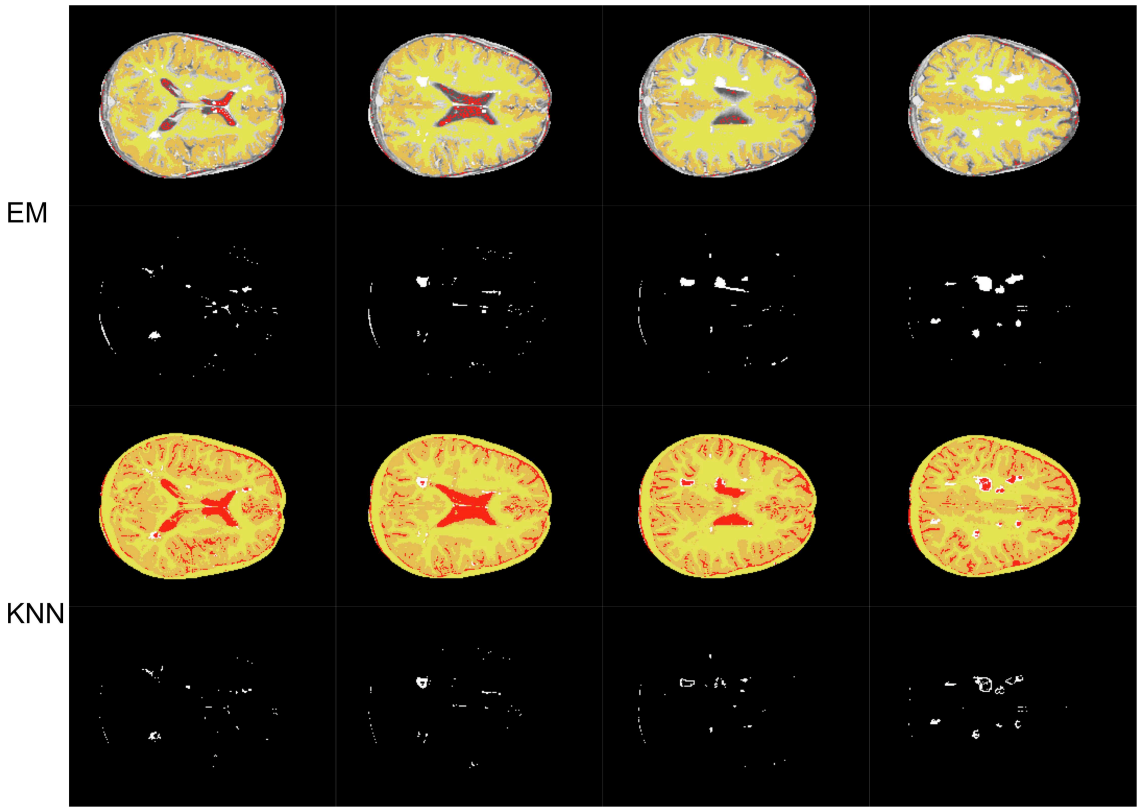


Figure 5.5: Real results using only affine transformations. Segmentations for the EM strategy (first and second row) and the kNN strategy (third and fourth row) for pure GM (orange), pure WM (yellow), pure CSF (red), PVE (different grey levels) and lesions (white) for different slices.

On the other hand, comparing the strategies according to local b-spline transformations as shown in figure 5.6, both methods present similar results for lesion segmentation. Moreover, some false positives are reduced thanks to a better delineation of tissues. However, the kNN strategy presents worse results with misclassification errors on the ventricular CSF, due to a better segmentation of external CSF.

5.3.2 Discussion

Both implemented methods are capable of segmenting lesions when a good alignment between the atlas and the image are provided. However, both strategies have some issues when dealing with tissues.

The approach from de Boer et al. [28] strongly depends on the samples provided. There-

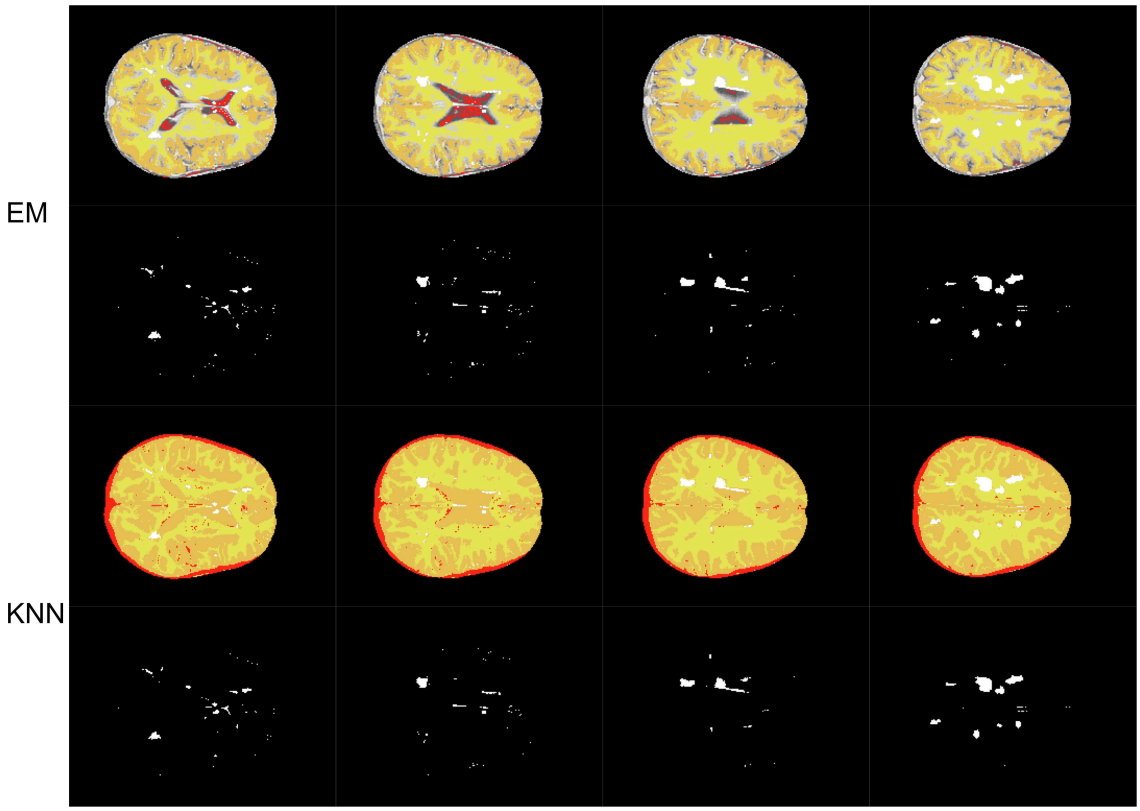


Figure 5.6: Real results using affine and b-splines transformations. Segmentations for the EM strategy (first and second row) and the kNN strategy (third and fourth row) for pure GM (orange), pure WM (yellow), pure CSF (red), PVE (different grey levels) and lesions (white) for different slices.

fore a bad registration (or only global) may produce a training set with mislabelled voxels. Those mislabelled voxels may vary the distribution of the training data set causing a misclassification of "apparent" GM voxels - according to intensity values - as CSF voxels modifying the computed threshold on the lesion segmentation step. Therefore, improving the registration reduces those issues. As a side effect, samples from ventricular CSF may not be taken into account reducing the accuracy on the CSF segmentation.

On the other hand, the approach from Souplet et al. [72], tries to correct PVE while using both the atlas information and intensity values. As a consequence, PVE classes contain most of the CSF voxels reducing the accuracy of the CSF segmentation as well. However, this statistical framework is more robust to atlas errors since Gaussian distributions are also taken into account reducing the bias caused by the registration.

5.4 Summary

We have tested our pipeline with both synthetic and real data sets. Evaluations were carried out quantitavely and qualitatively in order to point out strengths and weaknesses of our pipeline approach and our different tissues classification strategies. We have also seen the importance of a good registration in order to integrate atlas information. Global transformations can not capture all the differences between the template image and the new MRI scan, therefore, expected missalignments bias the segmentation process.

Comparing tissue strategies, we have seen how EM outperforms kNN using the Brainweb data. However, this images have a clear Gaussian distribution that previous intensity distribution estimation. Regarding real data, we have seen how both methods provide good tissue segmentations as input to our lesion segmentation step. However, a closer analysis of these masks presents missclassification errors in GM and CSF tissues that should be improved in order to use this information to compute volumetric measures (e.g. tissue atrophy).

Chapter 6

Conclusions and future work

After an extensive analysis of the state-of-the-art on MS lesion segmentation, the description of our proposal - including the implementation of the best recent strategies - and its final results, we extract our conclusions along with our contributions in this section. Moreover, we present our further work with a PhD thesis planning with further improvements on the methods presented.

6.1 Conclusions

Firstly, **we have reviewed the state-of-the-art on MS lesion segmentation**. After an extensive analysis of recent papers we have presented a new classification in supervised and unsupervised methods. Moreover, we divided supervised strategies between atlas-based and training-based, depending on how the a priori information is used; and we divided unsupervised strategies between those that use tissue information and those based only on lesion properties. Part of this work has been submitted to the IEEE Transactions on Information Technology in Biomedicine (TITB) journal and the 26th Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS).

Secondly, based on the conclusions of this state-of-the-art, **we decided to further review the state-of-the-art of atlas-based segmentation strategies of brain MRI**. According to the type of atlas used and how its information is used, we also presented a new classification of atlas-based segmentation strategies. This classification includes label propagation, multi-atlas fusion, feature definition and sample creation methods. An extensive review on this topic - partially included in this master thesis - has been submitted to the Computer Methods and Programs on Biomedicine (CMPB) journal.

Thirdly, **we designed and proposed a pipeline approach with preprocessing steps as well as registration and segmentation techniques**. Moreover, we chose two of the best recent methods on MS segmentation and we implemented them. Registration was implemented using the ITK library as a multiresolution framework including affine (global) and b-spline (local) transformations. Afterwards, atlas information was used in two different tissue segmentation strategies based on the EM and the kNN algorithms. The resulting tissue masks where then used to threshold the FLAIR image and obtain the lesion segmentation after a refinement step. It is important to note that our lesion segmentation approach is highly dependent on the FLAIR image and the tissue segmentation step.

Finally, we tested our proposal with both the Brainweb synthetic phantom and real data. Synthetic results were presented using quantitative measures (the DSC similarity index and the TPF), while the real results, due to the lack of ground truth, were evaluated qualitatively. We have seen how our proposal, which is an starting point for a PhD thesis, is capable of segmenting healthy tissues and MS lesions with similar results to other state-of-the-art approaches when using synthetic data. Moreover, we have seen that both strategies have similar results with lesions on real data, despite EM outperforming kNN with synthetic data.

6.2 Further work

In this master thesis we have presented a pipeline approach to segment MS lesions using atlas information. However, there are some improvements we could apply. For instance, as stated in chapter 5 some tissues have independent voxels labelled as tissue. Moreover, lesion masks present small false positives of sizes ranging between 1 and 3 unconnected voxels. Those missclassifications caused basically by noise can be avoided by applying stronger constraints to the segmentation. Those constraints can be applied using Markov random fields (MRF) theory. This stage will be developed during a research stay in the Ecole Polytechnique Federale de Lausanne (EPFL), in Switzerland.

As stated before, this master thesis is located within two funded projects in collaboration with 3 different hospitals, with different scanning machines. We plan on gathering 15 different temporal cases from each hospital to create a private MS database to further evaluate and validate our automatic segmentation algorithms. These cases will also be segmented by expert radiologists in order to provide ground truth information.

All these points, as well as possible new post-processing and pre-processing techniques

will be studied and developed as part of the PhD thesis. In the next section we present our planning for the PhD thesis.

6.2.1 Thesis planning

After submitting this master thesis we will continue this research line as part of the PhD thesis under a FI grant (reference 2010FI_B 00574) awarded by the Generalitat of Catalunya. Our main goal is the design an implementation of an automatic system to segment brain tissue and MS lesions in brain MRI. The framework presented in this thesis will be our starting point. A scheme of our planning to accomplish this goal is presented in figure 6.1. We will further study techniques to segment both abnormal and healthy tissues using a priori knowledge and MRI intensities in order to optimise our current approach.

For this purpose, we will study Markov random field methods in order to introduce stronger spatial constraints within our segmentation framework [53]. As stated before, those techniques will be developed during a research stay in the EPFL under a BE grant by the Generalitat de Catalunya. Starting on September of 2010 until April of 2011, this collaboration will extend the developed pipeline of this master thesis and improve segmentation results. Our prevision is to present those results in a conference and an international journal.

In order to further improve segmentation results, we also plan on further study registration algorithms as well as preprocessing tools. Those new techniques will be analysed during the PhD thesis and developed using the best state-of-the-art methods.

During the PhD thesis we will also keep gathering different MRI cases. Those cases will include healthy patients - without lesions - and MS patients. The goal of this recollection is to create both an atlas database, based on the healthy cases with manual or semi-automatic segmentations; and a MS evaluation database, with expert annotations for the lesions. We will study different methods to develop a probabilistic atlas based on those healthy cases to introduce a priori information on our framework. We also plan on publishing those atlas creation results on an international conference and journal.

Finally, we will integrate all the above-mentioned steps into our pipeline (MRF, new pre-processing and registration techniques and our own atlas) to create a better approach. This approach will also be tested against our final MS database as well as the MS Challenge 2008 database to compare our results with other state-of-the-art approaches.

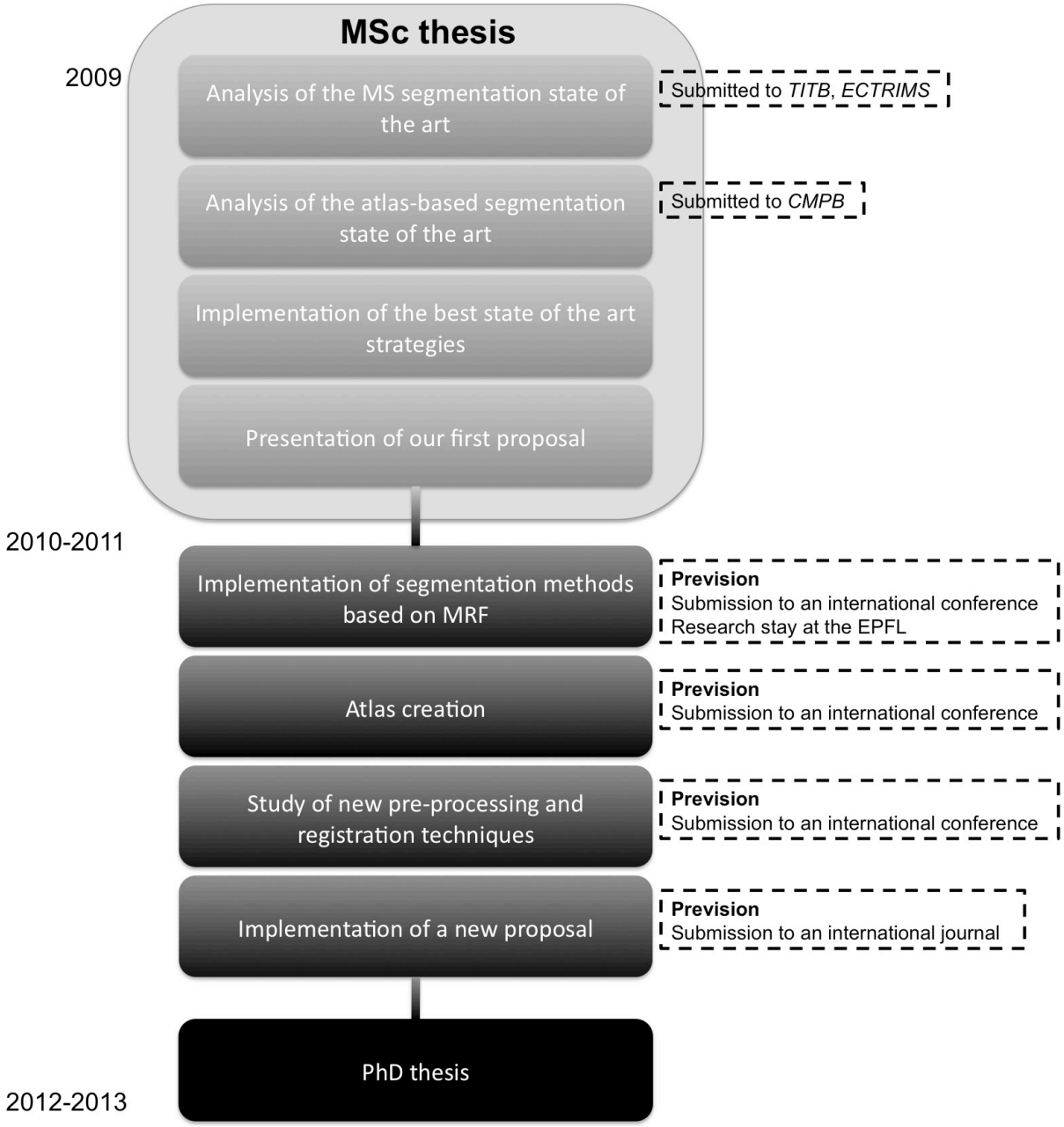


Figure 6.1: PhD planning scheme.

Bibliography

[1] A. Akselrod-Ballin, M. Galun, J.M. Gomori, M. Filippi, P. Valsasina, R. Basri, and A. Brandt. Automatic segmentation and classification of multiple sclerosis in multichannel MRI. *IEEE Trans. Biomed. Eng.*, 56(10):2461–2469, 2009.

[2] B. Alfano, A. Brunetti, M. Larobina, M. Quarantelli, E. Tedeschi, A. Ciarmiello, E. M. Covelli, and M. Salvatore. Automated segmentation and measurement of global white matter lesion volume in patients with multiple sclerosis. *J. Magn. Reson. Imag.*, 12(6):799–807, 2000.

[3] P. Anbeek, K. L. Vincken, G. S. van Bochove, M. J.P. van Osch, and J. van der Grond. Probabilistic segmentation of brain tissue in MR imaging. *NeuroImage*, 27(4):795–804, 2005.

[4] P. Anbeek, K. L. Vincken, M. J.P. van Osch, R. H.C. Bisschops, and J. van der Grond. Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage*, 21(3):1037–1044, 2004.

[5] P. Anbeek, K. L. Vincken, and M. A. Viergever. Automated MS-lesion segmentation by k-nearest neighbor classification. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

[6] A. Andronache, M. von Siebenthal, G. Szekely, and Ph. Cattin. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Med. Image Anal.*, 12(1):3 – 15, 2008.

[7] J. Ashburner and K. J. Friston. Nonlinear spatial normalization using basis functions. *Hum. Brain Mapp.*, 7(4):209 – 217, 1999.

[8] J. Ashburner and K. J. Friston. Voxel-based morphometry: the methods. *NeuroImage*, 11(6):805–821, 2000.

[9] L.S. Aït-Ali, S. Prima, P. Hellier, B. Carsin, G. Edan, , and C. Barillot. STREM: a robust multidimensional parametric method to segment MS lesions in MRI. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 409–416, 2005.

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

[11] P. L. Bazin and D. L. Pham. Homeomorphic brain image segmentation with topological and statistical atlases. *Med. Image Anal.*, 12(5):616 – 625, 2008.

[12] B. J. Bedell and P. A. Narayana. Automatic segmentation of gadolinium-enhanced multiple sclerosis lesions. *Magn. Reson. Med.*, 39(6):935–940, 1998.

[13] J. C. Bezdek. *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[14] J. C. Bezdek, L. O. Hall, and L. P. Clarke. Review of MR image segmentation techniques using pattern recognition. *Med. Phys.*, 20(4):1033–1048, 1993.

[15] K. Boesen, K. Rehm, K. Schaper, S. Stoltzner, R. Woods, E. Lï¿½ders, and D. Rottenberg. Quantitative comparison of four brain extraction algorithms. *NeuroImage*, 22(3):1255–1261, 2004.

[16] A.O. Boudraa, S. M. R. Dehak, Y.M. Zhu, C. Pachai, Y.G. Bao, and J. Grimaud. Automated segmentation of multiple sclerosis lesions in multispectral MR imaging using fuzzy clustering. *Comput. Biol. Med.*, 30(1):23–40, 2000.

[17] P.A. Brex, O. Ciccarelli, J.I. O'Riordan, M. Sailer, A.J. Thompson, and D.H. Miller. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *New Engl. J. Med.*, 346(3):158–164, 2002.

[18] S. Bricq, C. Collet, and J. P. Armspach. MS lesion segmentation based on hidde Markov chains. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

[19] S. Bricq, Ch. Collet, and J. P. Armspach. Unifying framework for multimodal brain MRI segmentation based on Hidden Markov Chains. *Med. Image Anal.*, 12(6):639 – 652, 2008.

[20] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger. MRI segmentation: Methods and applications. *Magn. Reson. Imag.*, 13(3):343–368, 1995.

[21] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans. A fully automatic and robust brain MRI tissue classification method. *Med. Image Anal.*, 7(4):513 – 527, 2003.

[22] D. L. Collins, A. P. Zijdenbos, V. Kollokian, and J. G. Sled. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imag.*, 17(3):463–468, 1998. http://www.bic.mni.mcgill.ca/brainweb.

[23] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(5):603 – 619, 2002.

[24] A. Compston and A. Coles. Multiple sclerosis. *Lancet*, 359(9313):1221–1231, 2002.

[25] R. Cárdenes, R. de Luis-García, and M. Bach-Cuadra. A multidimensional segmentation evaluation for medical image data. *Comput. Meth. Prog. Biomed.*, 96(2):108 – 124, 2009.

[26] S. Datta, B. R. Sajja, R. He, R. K. Gupta, J. S. Wolinsky, and P. A. Narayana. Segmentation of gadolinium-enhanced lesions on MRI in multiple sclerosis. *J. Magn. Reson. Imag.*, 25(5):932–937, 2007.

[27] S. Datta, B. R. Sajja, R. He, J. S. Wolinsky, R. K. Gupta, and P. A. Narayana. Segmentation and quantification of black holes in multiple sclerosis. *NeuroImage*, 29(2):467–474, 2006.

[28] R. de Boer, H. A. Vrooman, F. van der Lijn, M. W. Vernooij, M. A. Ikram, A. van der Lugt, M. M. B. Breteler, and W. J. Niessen. White matter lesion extension to automatic brain tissue segmentation on MRI. *NeuroImage*, 45(4):1151 – 1161, 2009.

[29] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. B*, 39(1):1 – 38, 1977.

[30] L. Devroye. On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Stat.*, 9(6):1310 – 1319, 1981.

[31] T. Dua and P. Rompani. *Atlas: Multiple sclerosis resources in the world 2008*. World Health Organization, 2008.

[32] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, 2 edition, 2001.

[33] G. Dugas-Phocion, M.A.G. Ballester, G. Malandain, C. Lebrun, and N. Ayache. Improved EM-based tissue segmentation and partial volume effect quantification in

multi-sequence brain MRI. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 26–33, 2004.

[34] F. Fazekas, F. Barkhof, M. Filippi, R. I. Grossman, D. K. B. Li, W. I. McDonald, H. F. McFarland, D. W. Paty, J. H. Simon, J. S. Wolinsky, and D. H. Miller. The contribution of magnetic resonance imaging to the diagnosis of multiple sclerosis. *Neurology*, 53(3):448–457, 1999.

[35] M. Filippi and F. Agosta. Imaging biomarkers in multiple sclerosis. *J. Magn. Reson. Imag.*, 31(4):770–788, 2010.

[36] M. Filippi, V. Dousset, H.F. McFarland, D.H. Miller, and R.I. Grossman. Role of magnetic resonance imaging in the diagnosis and monitoring of multiple sclerosis: Consensus report of the white matter study group. *J. Magn. Reson. Imag.*, 15(5):499–504, 2002.

[37] O. Freifeld, H. Greenspan, and J. Goldberger. Lesion detection in noisy MR brain images using constrained GMM and active contours. In *IEEE Int. Symp. Biomed. Imag.*, pages 596–599, 2007.

[38] D. García-Lorenzo, S. Prima, D. L. Collins, D. L. Arnold, S. P. Morrissey, and C. Barillot. Combining robust expectation maximization and mean shift algorithms for multiple sclerosis brain segmentation. In *Work. Med. Image Anal. Mult. Scler.*, pages 82–91, 2008.

[39] D. García-Lorenzo, S. Prima, S. P. Morrissey, and C. Barillot. A robust expectation-maximization algorithm for multiple sclerosis lesion segmentation. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

[40] D. García-Lorenzo, S. Prima, L. Parkes, J. C. Ferré, S. P. Morrissey, and C. Barillot. The impact of processing workflow in performance of automatic white matter lesion segmentation in multiple sclerosis. In *Work. Med. Image Anal. Mult. Scler.*, pages 104–112, 2008.

[41] D. Goldberg-Zimring, A. Achiron, S. Miron, M. Faibel, and H. Azhari. Automated detection and characterization of multiple sclerosis lesions in brain MR images. *Magn. Reson. Imag.*, 16(3):311–318, 1998.

[42] H. Greenspan, A. Ruf, and J. Goldberger. Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE Trans. Med. Imag.*, 25(9):1233–1245, 2006.

[43] S. W. Hartley, A. I. Scher, E. S. C. Korf, L. R. White, and L. J. Launer. Analysis and validation of automated skull stripping tools: A validation study based on 296 MR images from the Honolulu Asia aging study. *NeuroImage*, 30(4):1179–1186, 2006.

[44] R. He and P. A. Narayana. Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis. *Med. Phys.*, 29(7):1536–1546, 2002.

[45] P. Hellier and C. Barillot. A hierarchical parametric algorithm for deformable multimodal image registration. *Comput. Meth. Prog. Biomed.*, 75(2):107 – 115, 2004.

[46] Z. Hou. A review on MR image intensity inhomogeneity correction. *Int. J. Biomed. Imag.*, 2006(1):1–11, 2006.

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

[48] M. Kamber, R. Shinghal, D. L. Collins, G. S. Francis, and A. C. Evans. Model-based 3-D segmentation of multiple sclerosis in magnetic resonance brain images. *IEEE Trans. Med. Imag.*, 4(3):442–453, 1995.

[49] L. Kappos, D. Moeri, E. W. Radue, A. Schoetzau, K. Schweikert, F. Barkhof, D. Miller, C. R.G. Guttmann, H.L. Weiner, C. Gasperini, and M. Filippi. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. *Neurology*, 353(9157):964–969, 1999.

[50] R. Khayati, M. Vafadust, F. Towhidkhah, and S. M. Nabavi. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. *Comput. Biol. Med.*, 38(3):379–390, 2008.

[51] D.J. Kroon, E. van Oort, and K. Slump. Multiple sclerosis detection in multispectral magnetic resonance images with principal components analysis. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

[52] J. Lecoeur, J. C. Ferré, and C. Barillot. Optimized supervised segmentation of MS lesions from multispectral MRIs. In *Work. Med. Image Anal. Mult. Scler.*, pages 5–14, 2009.

[53] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans. Med. Imag.*, 20(8):677–688, 2001.

[54] M. Martin-Fernandez, E. Muñoz-Moreno, L. Cammoun, J.-P. Thiran, C.-F. Westin, and C. Alberola-López. Sequential anisotropic multichannel Wiener filtering with Rician bias correction applied to 3D regularization of DWI data. *Med. Image Anal.*, 13(1):19–35, 2009.

[55] W. I. McDonald, A. Compston, G. Edan, D. Goodkin, H.-P. Hartung, F. D. Lublin, H. F. McFarland, D. W. Paty, C. H. Polman, S. C. Reingold, M. Sandberg-Wollheim, W. Sibley, A. Thompson, S. Van Den Noort, B. Y. Weinshenker, and J. S. Wolinsky. Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann. Neurol.*, 50(4):121–127, 2001.

[56] H.F. McFarland, L.A. Stone, P.A. Calabresi, H Maloni, C.N. Bash, and J.A. Frank. MRI studies of multiple sclerosis: implications for the natural history of the disease and for monitoring effectiveness of experimental therapies. *Mult. Scler.*, 2(4):198–205, 1996.

[57] D. H. Miller, R. I. Grossman, S. C. Reingold, and H. F. McFarland. The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain*, 121(1):3–24, 1998.

[58] J. H. Morra, Z. Tu, A. W. Toga, and P. M. Thompson. Automatic segmentation of MS lesions using a contextual model for the MICCAI grand challenge. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

[59] V. Noblet, C. Heinrich, F. Heitz, and J. P. Armspach. 3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization. *IEEE Trans. Image Processing*, 14(5):553 – 566, 2005.

[60] G. Postelnicu, L. Zollei, and B. Fischl. Combined volumetric and surface registration. *IEEE Trans. Med. Imag.*, 28(4):508 – 522, 2009.

[61] M. Prastawa and G. Gerig. Automatic MS lesion segmentation by outlier detection and information theoretic region partitioning. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

[62] K. Rehm, K. Schaper, J. Anderson, R. Woods, S. Stoltzner, and D. Rottenberg. Putting our heads together: a consensus approach to brain/non-brain segmentation in T1-weighted MR volumes. *NeuroImage*, 22(3):1262–1270, 2004.

[63] A. Rovira, J.C.A. Cermeño, G.Izquierdo, and J.M.P.M. Tintoré. *Recomendaciones para la utilización de la RM en el diagnóstico y seguimiento de la Esclerosis Múltiple.* Grupo de enfermedades desmielizantes de la SEN, 2008.

[64] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imag.*, 18(8):712 – 721, 1999.

[65] B. R. Sajja, S. Datta, R. He, M. Mehta, R. K. Gupta, J. S. Wolinsky, and P. A. Narayana. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Ann. Biomed. Eng.*, 34(1):142–151, 2006.

[66] M. Scully, V. Magnotta, C. Gasparovic, P. Pelligrino, D. Feis, and H. J. Bockholt. 3D segmentation in the clinic: A grand challenge II at MICCAI 2008 - MS lesion segmentation. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

[67] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13(5):856–876, 2001.

[68] N. Shiee, P. L. Bazin, and D. L. Pham. Multiple sclerosis lesions segmentation using statistical and topological atlases. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

[69] N. Shiee, P.L. Bazin, J. L. Cuzzocreo, D. S. Reich, P. A. Calabresi, and D. L. Pham. Topologically constrained segmentation of brain images with multiple sclerosis lesions. In *Work. Med. Image Anal. Mult. Scler.*, pages 71–81, 2008.

[70] Navid Shiee, Pierre-Louis Bazin, Arzu Ozturk, Daniel S. Reich, Peter A. Calabresi, and Dzung L. Pham. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage*, 49(2):1524 – 1535, 2010.

[71] S. M. Smith. Fast robust automated brain extraction. *Hum. Brain Mapp.*, 17(3):143–155, 2002.

[72] J. C. Souplet, C. Lebrun, N. Ayache, and G. Malandain. An automatic segmentation of T2-FLAIR multiple sclerosis lesions. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, pages 1 – 11, 2008.

[73] J.C. Souplet, C. Lebrun, S. Chanalet, N. Ayache, and G. Malandain. Revue des approches de segmentation des lesions de sclerose en plaques dans les sequences conventionnelles IRM. *Revue Neurologique.*, 165(1):7–14, 2009.

[74] C.J. Stone. Consistent nonparametric regression. *Ann. Stat.*, 5(4):595 – 620, 1977.

[75] M. Styner, J. Lee, B. Chin, M.S. Chin, O. Commowick, H. Tran, V. Jewells, and S. Warfield. Editorial: 3D segmentation in the clinic: A grand challenge II: Ms lesion segmentation. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, pages 1–8, 2008.

[76] N.K. Subbanna, M. Shah, S. J. Francis, S. Narayannan, D. L. Collins, D. L. Arnold, and T. Arbel. MS lesion segmentation using Markov Random Fields. In *Work. Med. Image Anal. Mult. Scler.*, pages 15–26, 2009.

[77] J. P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med. Image Anal.*, 2(3):243 – 260, 1998.

[78] X. Tomas and S. K. Warfield. Fully-automatic generation of training points for automatic multiple sclerosis segmentation. In *Work. Med. Image Anal. Mult. Scler.*, pages 5–14, 2009.

[79] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imag.*, 18(10):897–908, 1999.

[80] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard. A level-set based approach to image registration. In *IEEE Work. Math. Meth. Biomed. Imag. Anal.*, pages 86 – 93, 2000.

[81] Baba C. Vemuri, Shuangying Huang, Sartaj Sahni, Christiana M. Leonard, Cecile Mohr, Robin Gilmore, and Jeffrey Fitzsimmons. An efficient motion estimator with application to medical image registration. *Med. Image Anal.*, 2(1):79 – 98, 1998.

[82] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(Supplement 1):S61 – S72, 2009.

[83] U. Vovk, F. Pernus, and B. Likar. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans. Med. Imag.*, 26(3):405–421, 2007.

[84] L. Wang, H. Lai, A. Thompson, and D. Miller. Survey of the distribution of lesion size in multiple sclerosis: implication for the measurement of total lesion load. *J. Neurol. Neurosurg. Psychiatry*, 63(4):452–455, 1997.

[85] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imag.*, 23(7):903 – 921, 2004.

[86] Y. Wu, S. K. Warfield, I. Leng Tan, W. M. Wells III, D. S. Meier, R. A. van Schijndel, F. Barkhof, and C. R. G. Guttmann. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. *NeuroImage*, 32(3):1205–1215, 2006.

[87] A. P. Zijdenbos, R. Forghani, and A. C. Evans. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis. *IEEE Trans. Med. Imag.*, 21(10):1280–1291, 2002.