# Contributions to Brain MRI Processing and Analysis

By

María Teresa García Sebastián

Dissertation presented to the Department of Computer Science and Artificial Intelligence in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

PhD Advisor:

Prof. Manuel Graña Romay

At
The University of the Basque Country
Donostia - San Sebastian
2009

AUTORIZACION DEL/LA DIRECTOR/A DE TESIS

PARA SU PRESENTACION

Dr/a. _____con N.I.F._____

como Director/a de la Tesis Doctoral: _____

_____

_____

realizada en el Departamento _____

_____

por el Doctorando Don/ña. _____ ,

autorizo la presentación de la citada Tesis Doctoral, dado que reúne las condiciones

necesarias para su defensa.

En _____a _____de _____de _____

EL/LA DIRECTOR/A DE LA TESIS

Fdo.: _____

CONFORMIDAD DEL DEPARTAMENTO

El Consejo del Departamento de _____

_____

en reunión celebrada el día _____ de _____ de _____ ha acordado dar la

conformidad a la admisión a trámite de presentación de la Tesis Doctoral titulada: _____

_____

_____

dirigida por el/la Dr/a. _____

_____

y presentada por Don/ña. _____

ante este Departamento.

En _____ a _____ de _____ de _____

Vº Bº DIRECTOR/A DEL DEPARTAMENTO   SECRETARIO/A DEL DEPARTAMENTO

Fdo.: _____   Fdo.: _____

ACTA DE GRADO DE DOCTOR

ACTA DE DEFENSA DE TESIS DOCTORAL

DOCTORANDO DON/ÑA. _____

TITULO DE LA TESIS: _____

_____

_____

El Tribunal designado por la Subcomisión de Doctorado de la UPV/EHU para calificar la Tesis Doctoral arriba indicada y reunido en el día de la fecha, una vez efectuada la defensa por el doctorando y contestadas las objeciones y/o sugerencias que se le han formulado, ha otorgado por_____la calificación de:

unanimidad ó mayoría

En _____a ____de _____de ____

EL/LA PRESIDENTE/A,   EL/LA SECRETARIO/A,

Fdo.:   Fdo.:

Dr/a: _____   Dr/a: _____

VOCAL 1º,   VOCAL 2º,   VOCAL 3º,

Fdo.:   Fdo.:   Fdo.:

Dr/a: _____   Dr/a:_____   Dr/a:_____

EL/LA DOCTORANDO/A,

Fdo.: __

Agradecimientos

Me gustaría agradecer a toda la gente que me ha ayudado, inspirado y sobre todo a las personas que me han apoyado moralmente durante estos cuatro últimos años, sin las que este trabajo no hubiera sido posible.

En especial quiero agradecer a mi director de Tesis, el Prof. Manuel Graña Romay, que sin su ánimo, guía y apoyo habría sido difícil superar ciertas situaciones ("*no pienses si eres o no eres capaz, simplemente hazlo*").

A mis amigas/os por su infinita paciencia y comprensión, pero sobre todo a Ana y Cristina que son las que más me han sufrido.

A mis amigos del grupo de Inteligencia Computacional y en especial a las personas que forman parte o colaboran en la línea de imagen médica.

Quiero agradecer de corazón el cariño de mis padres, hermanos y sobrinos, por estar siempre que los he necesitado. Y por último recordar a mi abuelo Victorino por la fuerza que me sigue dando.

¡Muchas gracias a todos!

Maite García.

Contributions to Brain MRI Processing and Analysis

by

María Teresa García Sebastián

abstract

This PhD Thesis work has been devoted to several aspects of the processing of medical images based on the Magnetic Resonance Imaging (MRI) technology. The works have been performed pursuing several lines of research that are somehow orthogonal. We started working on MRI denoising, removing intensity inhomogeneity effects. Then we have shifted our attention to the study of Alzheimer's disease and its detection on structural MRI (sMRI). We have explored the application of statistical and Artificial Neural Networks classifiers to feature vectors extracted in an specific way from the sMRI brain volumes. Finally, we have explored the application of novel computational algorithms proposed by our research group to the analysis of functional MRI (fMRI). These latter techniques fall in the category of Lattice Computing algorithms, and our aim was to show that they can, at least, perform as well as other statistical techniques for the detection of activation patterns.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter we will provide some introductory motivation remarks in section 1.1. In section 1.2 we make a quick summary of the publications produced during this PhD Thesis works. In section 1.3 we give a short relation of the research projects in which the candidate has been involved during the PhD Thesis works. Section 1.4 summarizes the contributions of this PhD Thesis work. In section 1.5 we comment the structure of this PhD Thesis report.

1.1 Motivation

Magnetic Resonance Imaging (MRI) is a medical imaging process, based on the phenomenon known as Nuclear Magnetic Resonance (NMR), which is gaining widespread aceptance for a large variety of medical explorations [46]. It is one of the most commonly used techniques in radiology to visualize the body soft tissues with great contrast, including the brain. Some reasons for this growing interest are:

- It is a noninvasive, based on non-ionizing radiation, imaging procedure.

- The fast evolution of MRI imaging techniques that offer a wide repertoire of pulse sequences that can easily be tuned to offer specific visualizations. MRI has become a very flexible imaging tool.

- The development of computer-assisted image processing tools including the quantification of the volumes of anatomical structures.

- MRI has a high spatial resolution and provides much information on the anatomical structure, allowing quantitative pathological or clinical studies, the derivation of digitized anatomical atlases and guidance before and during therapeutic intervention.

There are several different types of MRI procedures, but this PhD work is focused on structural MRI (sMRI) and functional MRI (fMRI), two of the most relevant techniques. The sMRI has a high spatial resolution and provides much information on the anatomical structure. On the other hand, fMRI allows to measure cerebral physiologic responses during neural activation and detect the brain areas which are involved in a task, a process or an emotion. This PhD Thesis work has three main avenues, each with its own focus and span. The first (chapter 2) is devoted to the correction of the intensity inhomogeneity (IIH) artifact in sMRI. The second (chapter 3) is devoted to the application of Machine Learning techniques to the detection of Alzheimer's disease (AD) in sMRI. The third (chapter 4) presents a novel approach to fMRI analysis. We review the motivation behind each chapter in the following subsections.

1.1.1 Intensity inhomogeneity (IIH) correction

The visualized MRI signal results from the aggregated measurements of the tissue composition at the molecular level. MRI images are expected to be piecewise constant except for partial volume effects in the tissue boundaries and the inevitable additive noise. However several imaging conditions introduce an additional multiplicative noise factor, referred to as the intensity inhomogeneity (IIH) field in the literature. The sources of IIH are generally divided in two groups [167]:

1. Related to properties of the MRI device such as static magnetic field inhomogeneity, RF signal energy spatial distribution and others.

2. Related to properties of the imaged object itself such as spatial variations of the magnetic permeability and dielectric properties of the imaged object.

The correction of the IIH artifact is the process of estimating the IIH field and removing (dividing or subtracting) this estimation from the given image. It is of interest in order to improve the performance of MRI segmentation algorithms. Under ideal conditions (no partial volume effect, no IIH, no variations due to the imaging device) the segmentation of MRI would be trivial once the signal intensity for each kind of tissue is known. However, the mentioned artifacts impose

the need for adaptive and robust segmentation algorithms. One of the abilities required is the estimation of the IIH field, whether it is due to partial volume effects or to a smooth spatial variation of the signal measurement. Some algorithms perform both the IIH field estimation and the MRI segmentation into the tissue regions of interest. This segmentation may be explicit or implicit, giving the estimation of the intensity class means corresponding to the imaged tissues.

A broad taxonomy of MRI IIH correction algorithms divides them between parametric and non-parametric algorithms. The first ones impose a parametric model of the inhomogeneity field [26, 111, 157, 65], whose parameter values must be estimated to fit the model. The non parametric algorithms [18, 81, 107, 111, 127, 149, 148, 168, 178] do not propose any model, so that they perform a non-parametric estimation of the inhomogeneity field value at each voxel position of the measured MRI volume. In this PhD Thesis works, we have been working on two different IIH correction approaches which have complementary advantages:

- A parametric algorithm (GradClassLeg) whose assumed model of the IIH field is a linear combinations of outer products of Legendre polynomials. Estimation is performed by error function gradient descent over both the inhomogeneity field model parameters and the tissue class intensity means. Therefore this can be an IIH correction and segmentation algorithm.

- A non-parametric approach (AFR), in which the value of the IIH field is estimated independently at each voxel site, using neighboring information which is similar to that of the Self Organizing Map (SOM).

1.1.2 Alzheimer's disease detection

During the last decade, interest in MRI for the study and diagnostic support of neurodegenerative diseases has been rapidly rising. The increase in studies of neurodegenerative diseases are due to the increase in the incidence and prevalence of these diseases. Alzheimer's disease (AD) is a neurodegenerative disorder, which is one of the most common cause of dementia in old people, with a high socio-economic impact in occidental countries. The diagnosis of AD is done after the exclusion of other forms of dementia but definitive diagnosis can only be made after a post-mortem study of the brain tissue. This is one of the reasons why MRI based early diagnosis is a current research hot topic in the neurosciences. The pharmaceutical companies have already recognized that imaging techniques especially MRI and Positron Emission Tomography (PET) provide "surrogate" information concerning the pattern and rate of neurodegeneration, which can be used

to monitor the effects of treatments which slow the progression of neurodegeneration. Therefore, there is high interest in the development of automated detection procedures based on MRI and other medical imaging techniques.

There are already studies about discrimination of AD patients from controls based on PET and Single-Photon Emission Tomography (SPECT) functional volumes [128, 140, 64, 19] applying Support Vector Machines (SVM). The works on MRI data are few, some applying Artificial Neural Networks [47, 86]. A critical aspect of these studies is the feature vector extraction process. The literature contains approaches ranging from the direct use of image intensities, subsampling and aggregation, Principal Component Analysis (PCA) transformations and the application of morphometry techniques [43, 53, 177, 86], such as Voxel Based Morphometry (VBM), to identify potentially discriminant voxels. Another critical aspect of the studies is the subject population, which is not well documented most of the times or correspond to in-house non publicly available collections.

We have performed an study over a carefully selected collection of images, selected from the OASIS database described in Appendix C. We have tested several Machine Learning classifier construction systems over feature vectors computed on the clusters of significant vectors detected by a VBM study. This study can be considered a state of the art contribution to the issue of automated detection of AD from sMRI.

1.1.3 fMRI analysis

Human brain mapping is a rapidly expanding discipline, and in recent years interest has grown in novel methods for imaging human brain functionality. Noninvasive techniques can measure cerebral physiologic responses during neural activation. One of the relevant techniques is functional Magnetic Resonance Imaging (fMRI) [122], which uses the blood oxygenation level dependent (BOLD) contrast. Since these methods are completely noninvasive, using no contrast agent or ionizing radiation, repeated single-subject studies are becoming feasible [117].

The fMRI experiment consists of a functional template or protocol (e.g., alternating activation and rest for a certain time) that induces a functional response in the brain. The aim of an fMRI experiment is to detect responses to this stimulus in identified voxels. The two main experimental designs used in fMRI are capable of effectively addressing two different questions of interest. Block designs are especially useful for detection, that is, locating which voxels are activated in response to a given task, compared to a control condition. Event-related designs, on the other hand, provide a means of estimating the hemodynamic response function.

The functional information of a voxel has to be extracted from its functional time course. Therefore, for each functional time point one fMRI volume is recorded. The complete four-dimensional dataset (three dimensions in space, one dimension in time) consists of subsequently recorded three-dimensional (3-D) volumes and thus for each voxel of a volume a functional time course exists. The acquisition of these functional volumes runs over periods lasting up to several minutes.

The most extended analysis approach for fMRI signals is the Statistical Parametric Map (SPM) [63]. There have been also approaches to the fMRI analysis based on the Independent Component Analysis (ICA) [30, 35, 34, 33, 80] assuming that the time series observations are linear mixtures of independent sources which can not be observed. In this PhD Thesis work we propose an innovative analysis procedure that combines Lattice Computing [77] techniques for endmember induction, and the unconstrained least squares estimation of the contribution of these endmembers to the observed time series at each voxel.

1.2 Publications produced during the PhD Thesis works.

Related to the intensity inhomogeneity correction (IIH) algorithms:

1. Maite García-Sebastián, Ana Isabel González, Manuel Graña. An Adaptive Field Rule for Non-parametric MRI Intensity Inhomogeneity Estimation Algorithm. *Neurocomputing*, 72, pp: 3556–3569, 2009.

2. Maite García-Sebastián, Carmen Hernández, Alicia d'Anjou. Robustness of an Adaptive MRI Segmentation Algorithm Parametric Intensity Inhomogeneity Modeling. *Neurocomputing,* 72, pp: 2146-2152, 2009.

3. Maite García-Sebastián, Elsa Fernández, Manuel Graña, Francisco J. Torrealdea. A Parametric Gradient Descent MRI Intensity Inhomogeneity Correction Algorithm. *Pattern Recognition Letters,* 28(13):1657-1666, 2007.

4. Maite García-Sebastián, Manuel Graña. SOM for Intensity Inhomogeneity Correction in MRI. *European Symposium on Artificial Neural Networks (ESANN)*, M. Verleysen (ed), D-side Publications, pp: 109-114, 2007. ISBN 2-930307-07-2.

5. Maite García-Sebastián, Ana Isabel González, Manuel Graña. Derivation of SOM-like Rules for Intensity Inhomogeneity Correction in MRI. *Computational and Ambient Intelligence (IWANN)*, F. Sandoval, A. Prieto, J. Cabestany, M. Graña (eds). LNCS 4507, Springer Verlag, Heidelberg, pp: 669-676, 2007. ISBN 3-540-73006-0.

6. Maite García-Sebastián, Alexandre M. Savio, Manuel Graña. Comments on an Evolutionary Intensity Inhomogeneity Correction Algorithm. *Proceedings CEC 2008*, pp: 4146-4150. ISBN 978-1-4244-1822-0.

7. Ana Isabel González, Alicia D'Anjou, Maite García-Sebastián, Manuel Graña. SOM and Neural Gas as Graduated Nonconvexity Algorithms. *The 2006 International Conference on Computational Science and its Applications (ICCSA)*, M. Gavrilova et al.(eds.), LNCS 3982, Springer Verlag, Berlin Heidelberg, part III, pp: 1143-1152, 2006. ISBN 3-540-34075-0.

8. Carmen Hernández, Josune Gallego, Maite García-Sebastián, Manuel Graña. On Clustering Performance Indices for Multispectral Image. *Knowledge-Based Intelligent Information and Engineering Systems (KES)*, B. Gabrys, R. J. Howlett and L. C. Jain (eds.), LNCS 4253, Springer Verlag, Berlin, Part III, pp: 277-283, 2006. ISBN 978-3-540-46542-3.

9. Maite García-Sebastián, Elsa Fernández, Manuel Graña, F.J. Torrealdea. A Gradient Descent MRI Illumination Correction Algorithm. *Computational Intelligence an Bioinspired Systems (IWANN)*, F. Sandoval, A. Prieto, J. Cabestany (eds), LNCS 3512, Springer Verlag, Berlin, vol. 3512, pp: 913-920, 2005. ISBN 3-540-26208-3.

Related to Alzheimer's disease detection by Machine Learning approaches based on structural MRI

1. Maite García-Sebastián, Alexandre M. Savio, Manuel Graña, Jorge Villanúa. On the Use of Morphometry Based Features for Alzheimer's Disease Detection on MRI. *Bio-Inspired Systems: Computational and Ambient Intelligence. / IWANN 2009 (Part I)*. Joan Cabestany, Francisco Sandoval, Alberto Prieto, Juan M. Corchado (Editors), LNCS 5517, Springer Verlag 2009, pp: 957-964. ISBN 978-3-642-02477-1.

2. Alexandre M. Savio, Maite García-Sebastián, Manuel Graña, Jorge Villanúa. Results of an Adaboost Approach on Alzheimer's Disease Detection on MRI. J. Mira, J. M. Ferrández, J.R. Alvarez, F. dela Paz, F.J. Toledo

(Eds.) Bioinspired Applications in Artificial and Natural Computation, *Proc. IWINAC 2009*, Part II, LNCS 5602, Springer Verlag, pp: 114-123. ISBN 978-3-642-02266-1.

3. Alexandre M. Savio, Maite García-Sebastián, Carmen Hernández, Manuel Graña, Jorge Villanúa. Classification Results of Artificial Neural Networks for Alzheimer's Disease Detection. *Proc IDEAL 2009*. IISBN pending

Related to the Lattice Computing approach for fMRI analysis

1. Manuel Graña, Alexandre M. Savio, Maite García-Sebastián, Elsa Fernández. A Lattice Computing approach for On-line fMRI analysis, *Image and Vision Computing* (accepted with minor revision)

2. Manuel Graña, Maite García-Sebastián, Carmen Hernández , Lattice Independent Component Analysis for fMRI, *Journal Information Sciences* (submitted)

3. Manuel Graña, Maite García-Sebastián, Carmen Hernández. Lattice Independent Component Analysis for fMRI analysis. M. Polycarpou, C. Panayiotou, C. Alippi, G. Ellinas (Eds.), Artificial Neural Networks - ICANN 2009, LNCS , Springer Verlag, in press.

4. Manuel Graña, Maite García-Sebastián, Ivan Villaverde, Elsa Fernández. An Approach from Lattice Computing to fMRI Analysis. LBM 2008 (CLA 2008), Proceedings of the Lattice-Based Modeling Workshop, Vassilis Kaburlasos, Uta Priss, Manuel Graña (eds), pp:33-44. ISBN 978–80–244–2112–4.

1.3 Research projects

During the PhD Thesis works, the candidate has been part of several research projects, some of them with indirect relation to the main topics of the PhD Thesis, but always related to image processing.

1. Desarrollo de métodos y técnicas de análisis de imagen de Resonancia Magnética cerebral en el ámbito del software libre para estudio de enfermedades neurodegenerativas (BRAINER), *programa SAIOTEK 2008-2009*

Eusko Jaurlaritza/Gobierno Vasco. Investigadora principal: Carmen Hernández Gómez. Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU), Fundación Vasca de Innovación e Investigación Sanitaria/B+I+O eusko fundazioa, Departamento de Neurociencia (UPV/EHU). Duración (meses): 24 meses. Referencia: SA-2008/00226.

2. Aportaciones al diagnóstico de Esclerosis Múltiple mediante el análisis de imagen de Resonancia Magnética, *programa "Red guipuzcoana de Ciencia, Tecnología e Innovación" Gipuzkoako Foru Aldundia/Diputación Foral de Guipúzcoa, 2006.* Investigadora principal: Elsa Fernández Gómez de Segura. Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU). Duración (meses): 24 meses. Referencia: OF0107/2006.

3. Percepción artificial y control de caos para robótica modular en entornos dinámicos y no estructurados, *Ministerio de Educación y Ciencia, 2006.* Investigador Principal: Manuel Graña Romay. Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU), Universidad de A Coruña, Inteligencia Artificial (UPM). Duración (meses): 36 meses. Referencia: DPI2006-15346-C03-03.

4. Computación distribuida para servicios Web multimedia: servidor de imágenes de múltiples modalidades, *programa SAIOTEK 2004-2005 Eusko Jaurlaritza/Gobierno Vasco.* Investigador principal: Manuel Graña Romay. Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU). Duración (meses): 24 meses. Referencia: OD04UN81.

5. Servidor de imágenes basado en contenidos para el control de calidad de papel reciclado basado en índices de textura, *proyectos Universidad-Empresa (UPV/EHU), 2004*. Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU), Recypel S.A.. Investigador principal: Manuel Graña Romay. Duración (meses): 24 meses. Referencia: UE03A07.

6. Métodos computacionales para la detección de vertidos a partir de información de sensores remotos hiperespectrales y otros, *Ministerio de Ciencia y Tecnología, 2003.* Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU), Universidad de A Coruña, Universidad de Santiago de Compostela, Universidad de Vigo. Investigador principal: Manuel Graña Romay. Duración (meses): 36 meses. Referencia: VEM2003-20088-C04-04.

7. Control de calidad del papel reciclado mediante visión artificial. *Ministerio de Ciencia y Tecnología, 2003.* Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU). Investigador principal: Manuel Graña Romay. Duración (meses): 36 meses. Referencia: DPI2003-06972.

The candidate has been also allowed by the research group support to propose and manage the following projects during the PhD Thesis works,

1. Desarrollo de técnicas de análisis de imágenes de Resonancia Magnética con aplicación en el diagnóstico preclínico de la enfermedad de Alzheimer, *programa SAIOTEK 2007-2008 Eusko Jaurlaritza/Gobierno Vasco. Investigadora principal: Maite García-Sebastián.* Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU). Duración (meses): 24 meses. Referencia: S-PR07UN02.

2. Detección preclínica del Alzheimer mediante Neuroimagen, *programa "Red guipuzcoana de Ciencia, Tecnología e Innovación" Gipuzkoako Foru Aldundia/Diputación Foral de Guipúzcoa. Investigadora principal: Maite García-Sebastián.* Entidades: Ciencias de la Computación e Inteligencia Artificial (UPV/EHU). Duración (meses): 12 meses. Referencia: OF145/2007.

1.4 Contributions of the PhD Thesis

We summarize here the contributions of this PhD Thesis work to the state of the art. Due to the fast pace of science, some of these contributions will be soon superseded by our own or another work, nevertheless we hope that some bits will remain useful for a time.

- We have performed an exhaustive review of the state of the art in several aspects of the processing of MRI related data. These reviews appear in the diverse chapters and annexes of this PhD report. The areas covered are:

  - Intensity inhomogeneity (IIH) correction and segmentation algorithms.
  - Classification approaches for early detection of AD based on MRI and other medical imaging techniques.
  - Applications of Independent Component Analysis (ICA) to fMRI data analysis.

- – Image resources available for the realization of computational experiments

- We have proposed a parametric IIH correction algorithm proving its convergence properties and testing it on well known benchmark data against state of the art algorithms. We have also done a thoroughful sensitivity analysis to asses its robustness against bad initializations.

- We have proposed a non-parametric IIH algorithm, proving its convergence properties and testing it on well known benchmark data. We have studied its interaction with several intensity class means estimation approaches, and we have detected a potential flaw in the segmentation of one of the publicised IBSR collection of MRI brain volumes.

- We have proposed a feature extraction methodology and tested several classification approaches to support the clinical diagnosis of Alzheimer's disease. The computational validation has been carried out on a subset of the OASIS database after a careful selection of the individuals involved in the experiment.

- We have proposed and tested a Lattice Independent Component Analysis (LICA) procedure for fMRI data analysis that consists of an endmember extraction algorithm that is based on Lattice Associative Memories and Lattice Independence properties. The approach performs comparable to ICA on a well known benchmark dataset.

1.5 Structure of the PhD Thesis report

The PhD Thesis report is structured along the three main avenues of work followed by the PhD candidate during the past four years. Figure 1.1 shows as a summary the relationship of the works and chapters. There are three chapters devoted to the main works and three appendices that provide additional information that may be useful to understand some issues or provide material for further works. The contents of the chapters in this PhD Thesis report are as follows:

- Chapter 2 is devoted to intensity inhomogeneity (IIH) correction, where we present the two algorithms which have been proposed and tested. We provide a description of the algorithms, the state of the art and antecedents of
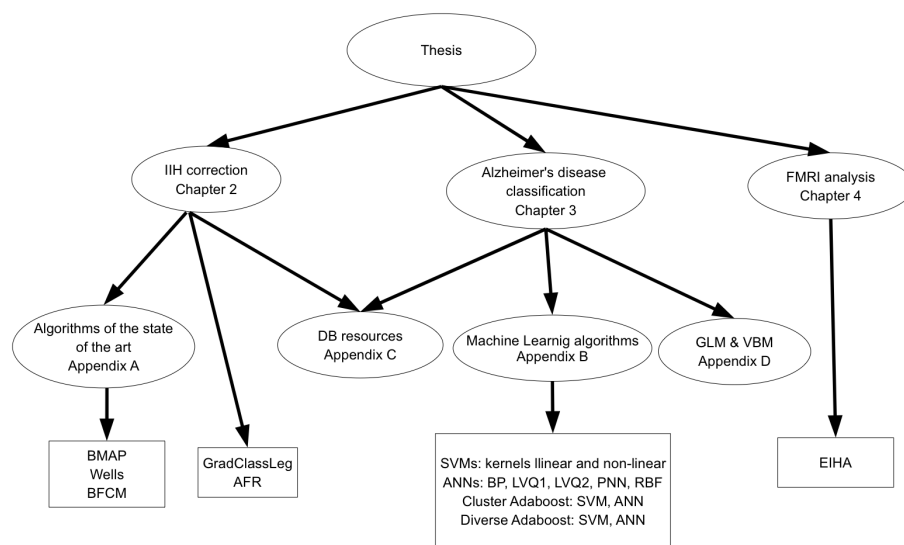
Figure 1.1: Structure of the PhD Thesis report
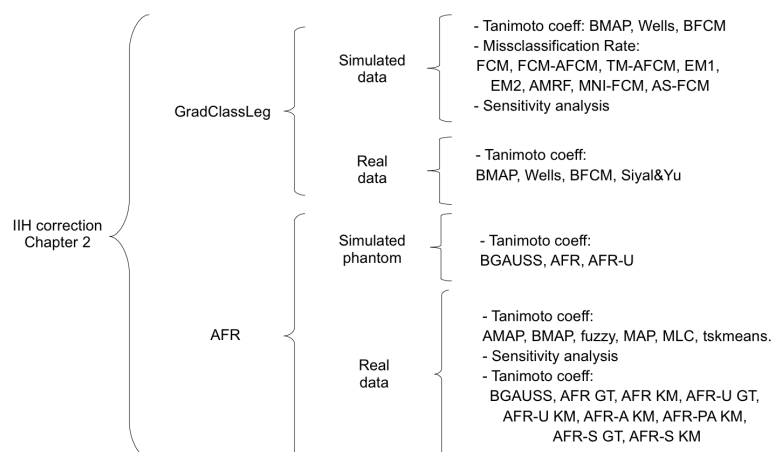
Figure 1.2: Summary of experiments in chapter 2
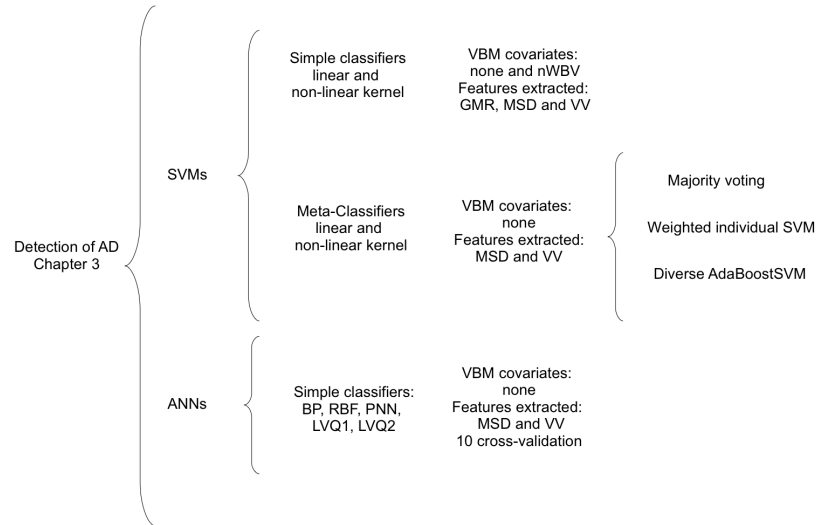
Figure 1.3: Summary of experiments in chapter 3

our proposals. We also report some computational experiments comparing our algorithms with other state of the art algorithms and performing exhaustive sensitivity analyses. Figure 1.2 shows a summary of the experiments presented in this chapter. For each of the proposed algorithms (GradClassLeg and AFR), figure 1.2 shows the kind of data used (simulated and real), the algorithms that were used as benchmark and the performance measure of the algorithm responses.

- Chapter 3 is devoted to the approaches pursued for detection of Alzheimer's disease (AD) based on sMRI. We present a feature extraction process that uses the well known Voxel Based Morphometry (VBM) in order to select the most significative voxels. Several feature vector extraction procedures are defined and tested, as well as several computational intelligence classification techniques, based on statistical classification and on Neural Network approaches. Figure 1.3 shows a summary of the experiments presented in this chapter for the early detection of AD, giving the combinations of Machine Learning approaches for the construction of the classification systems and feature extraction processes.

- Chapter 4 is devoted to the application of an endmember extraction algorithm based on Lattice Associative Memories to the task of activation detection in fMRI data. The fundamental results upon which our work relies are reviewed, the endmember induction algorithm is presented and some results on a well known case study are given.

- Appendix A gathers the formal definitions of some IIH correction algorithms that we use to compare to our proposals in chapter 2.

- Appendix B gathers the formal definitions of the Machine Learning algorithms used in chapter 3.

- Appendix C contains a review of the diverse database resources that can be found on Internet, some of which have been extensively used in our works.

- Appendix D contains a review of the General Linear Model and Statistical Parametric Map that complements the material in chapter 3.

1.5.1 About the conclusions sections

Each of the main chapters in this PhD Thesis report represents a different track of research, ending with a particular conclusions section. Therefore, there is no specific chapter devoted to conclusions for the whole report.

Chapter 2

MRI intensity inhomogeneity correction

In this chapter we present two algorithms for intensity inhomogeneity (IIH) correction in MRI that have been developed in the framework of this PhD Thesis works. These algorithms follow two divergent approaches and they have complementary advantages and disadvantages, as will be clarified below for each of them. We did not compare them among themselves, because they correspond to different paradigms and answer different needs separately, although we compare them to state of the art approaches found in the literature. The first one is a parametric algorithm, that we call GradClassLeg because the IIH field is modelled as a product of Legendre polynomials. The second one is a non-parametric algorithm inspired in the Self-Organized Map which we call Adaptive Field Rule (AFR).

The structure of the chapter is as follows: Section 2.1 gives some background information about our works, both regarding IIH correction algorithms and application of Self-Organizing Maps (SOM) to MRI analysis. Section 2.2 gives the formal definition of our algorithms (the formal definition of other state of the art algorithms are in Appendix A) including some discussions about convergence and other formal issues. Section 2.3 describes the computational experimentation elements which are common to the validation of both GradClassLeg and AFR algorithms. Section 2.4 reports the computational experiments performed to validate and explore the sensitivity of GradClassLeg. Section 2.5 reports the computational experiments performed to validate the AFR algorithm, and its interaction with different intensity class means estimation processes. Finally, section 2.6 gives separate conclusions for GradClassLeg and AFR.

2.1 Antecedents of our work

There are two kind of antecedents that we found relevant for the PhD Thesis works reported in this chapter. First is the state of the art of algorithms that perform IIH correction and segmentation of MRI data. The second is the previous uses of SOM for MRI data analysis and segmentation, and how our work departs from other approaches.

2.1.1 IIH correction algorithms

MRI allows to visualize with great contrast the soft tissues in the body and has revolutionized the capacity to diagnose the pathologies that affect them [46]. MRI has a high spatial resolution and provides much information on the anatomical structure, allowing quantitative pathological or clinical studies, the derivation of digitized anatomical atlases and a guidance before and during therapeutic intervention. It is based on the phenomenon known as Nuclear Magnetic Resonance (NMR). The visualized signal results from the aggregated measurements of the tissue composition at the molecular level. MRI images are expected to be piecewise constant except for partial volume effects in the tissue boundaries and the inevitable additive noise. Thus, once the expected intensities of each tissue are known, we could construct a good approximation to the optimal Bayesian classifier of minimum classification error, assuming that the intensity probability distribution is a mixture of Gaussians whose means are the tissue expected intensities. Then, we could apply this classifier to perform the image segmentation task. However several imaging conditions introduce an additional multiplicative noise factor, referred to as the intensity inhomogeneity (IIH) field in the literature. The sources of IIH are generally divided in two groups [167]:

1. Related to properties of the MRI device such as static magnetic field inhomogeneity, RF signal energy spatial distribution and others.

2. Related to the imaged object itself such as the specific magnetic permeability and dielectric properties of the imaged object.

Conventional clustering algorithms [49] can cope with the additive noise, but the multiplicative inhomogeneity field has catastrophic effects on them. A general algorithm for IIH correction is the lowpass filtering in the log-domain, which is equivalent to homomorphic filtering [76] in digital image processing for the correction of illumination inhomogeneity. However, it is of no use for MRI because

there is a great overlapping of the Fourier spectra corresponding to the IIH field and the image. Most IIH correction algorithms in MRI are composed of a method for IIH field estimation and a classification algorithm applied to the restored image obtained removing the IIH field. In some algorithms, the classification and bias estimation steps are interleaved in the iterations. In others, like the gradient descent algorithm GradClassLeg proposed below in section 2.2.1, these steps are performed simultaneously.

A broad taxonomy of MRI IIH correction algorithms divides them between parametric and non-parametric algorithms. The first ones impose a parametric model of the IIH field [26, 111, 157, 65], whose parameter values must be estimated to fit the model. The non parametric algorithms [18, 81, 107, 111, 127, 149, 148, 168, 178] do not propose any model, so that they perform a non-parametric estimation of the IIH field value at each voxel position of the measured MRI volume. A similar taxonomy can be found in [85]. We have already made a short review in [65].

From the point of view of the estimation procedure, the two approaches which have produced the greater number of works devoted to this issue are Bayesian image processing algorithms and Fuzzy Clustering. Bayesian algorithms [81, 111, 127, 168, 178] perform the Maximum A Posteriori (MAP) estimation of either the IIH field or the classification image, or both. The approach needs the formulation of an *a priori* model of the images and/or the IIH field probability density, and the conditional probability density of the observed image. The probability density of the observed image conditioned to the voxel classification and the IIH field is usually assumed to be Gaussian. The *a priori* model of the MRI images is sometimes specified by a Markov Random Field (MRF) that formalizes the smoothness constraints on the image classification [111, 127]. In [81, 168] modelling the bias as a Gaussian distributed random vector leads to the Expectation Maximization estimation algorithm of the IIH field. The Hidden Markov Random Field proposed in [178] is essentially identical to the MRF *a priori* density model of [127].

Fuzzy Clustering algorithms [18, 107, 148] perform the estimation of the image classification minimizing an objective function given by the voxel quantization error weighted by the fuzzy membership coefficients. The algorithms estimate the membership coefficients, the intensity class means and the IIH bias through this minimization process.

In this framework, the two approaches we have been working on in this PhD Thesis works can be characterized as follows:

- GradClassLeg is a gradient descent algorithm of the classification error

function of the image corrected with a parametric estimate of a IIH field, approximated by linear combinations of products of Legendre polynomials. The error function gradient descent is performed over both the IIH field model parameters and the tissue intensity class means. The relative advantages of GradClassLeg are the following:

– Works with a compact representation of the IIH field.

– The algorithm is fast because of the low dimension representation of the IIH field and because it is a gradient descent algorithm.

– The approach provides a global model for the IIH field, so that the estimation is non-local, less sensitive to local noise, and the estimated field is intrinsically smooth.

– The approach is robust against bad initializations of the intensity class means and the IIH field parameters.

• The Adaptive Field Rule (AFR) is a non-parametric approach, where the value of the IIH field is estimated independently at each voxel site. It could be easily programmed to run in parallel multiprocessor machines. The type of neighboring information used is similar to that of the Self Organizing Map (SOM), because the estimation of the IIH field at each voxel site is changed when its value at the neighboring sites is updated. The relative advantages of AFR are the following:

– The IIH field is represented at voxel level, thus AFR is able to model local features, such as partial volume effects.

– High robustness against bad intensity class means estimations.

2.1.2 SOM for MRI processing

In early applications of the SOM to MRI data [25, 160], it was used to estimate the clustering of intensity into several classes, either in multispectral or single modality images. In these works, the existence of IIH fields is not taken into account. The results that these works report show an overestimation of the number of classes in the image. This is a natural effect when the IIH field is not considered. More recent works use the SOM as a preprocessing step previous to manual examination [72, 119, 164] or the supervised construction of classification systems for image segmentation. In this sense SOM has been used to obtain the reference

vectors for the construction of a supervised Probabilistic Neural Network (PNN) [152], a multi-layer feedforward neural network with automated Bayesian regularization [88], Support Vector Machine (SVM) [37]. Recent applications include the detection of activity patterns in fMRI [100] which are too low resolution to try to detect IIH effects.

The main difference between the AFR, presented in this chapter in section 2.2.3, and the previous SOM approaches to MRI processing, is that we focus on the estimation of the IIH field. This assumption is grounded in a widely accepted imaging model presented in the next section. When we propose adaptive algorithms for the estimation of the intensity class means, it is done in the precise context of the tissues we are looking for (i.e., Cerebrospinal fluid (CSF), gray matter (GM), white matter (WM)).

2.2 Description of the algorithms

We will denote $\mathbf{y} = (y_i; i \in I)$ the observed image and $\mathbf{x} = (x_i; i \in I, \ x_i \in \Omega)$ the underlying tissue classification image, where $i \in I \subset N^3$ is the voxel site in the discrete lattice of the image support for 3D images, and $\Omega = \{\omega_1, ..., \omega_c\}$ is the set of tissue classes in the image. The assumed image formation model is the following one:

$$y_i = \beta_i \cdot r_i + \eta_i, \tag{2.1}$$

where β_i is the multiplicative IIH field, r_i is the clean signal associated with the true voxel class x_i and η_i is the additive noise. In MRI we have the additional restriction that the signal intensity values belong to a discrete (small) set, $\Gamma = \{\mu_{\omega_1}, ..., \mu_{\omega_c}\}$, so that $r_i = \mu_{x_i}$.

Definition The IIH *robust segmentation problem* is the problem of estimating the image segmentation $\mathbf{x}$ and the inhomogeneity multiplicative field $\beta = (\beta_i; i \in I)$ from $\mathbf{y}$.

Definition The IIH *correction problem* is that of estimating the inhomogeneity multiplicative field $\beta = (\beta_i; i \in I)$ from $\mathbf{y}$ and computing the corrected image:

$$\hat{y}_i = \frac{y_i}{\beta_i} \tag{2.2}$$

Both problems are ill posed, and closely related. We have found that [149] is the only work that performs an IIH correction without resorting to image segmentation. In our works we sometimes assume that the image segmentation is given through some preset intensity class means. However both approaches (GradClassLeg and AFR) are formulated allowing the possibility to perform the estimation of the image segmentation.

In some algorithms the correction and estimation is performed on the image logarithm. If we discard the additive noise term, we have that the image formation model described in equation (2.1) becomes:

$$Y_i = B_i + R_i,$$ (2.3)

where $Y_i = \ln y_i$, $B_i = \ln \beta_i$ and $R_i = \ln r_i$. The log-images are denoted as $\mathbf{Y} = (Y_i; i \in I)$, $\mathbf{B} = (B_i; i \in I)$ and $\mathbf{R} = (R_i; i \in I)$. The multiplicative field β becomes an additive term $\mathbf{B}$ and, because of that, it is usually named the *bias* field. We maintain the definition of the tissue classes and the corresponding intensity class means, so that $R_i = M_{x_i}$, where $M_\omega = \ln \mu_\omega$.

The shift between the two image formation models is not as trivial as may appear at first sight, because the log-model in equation (2.3) implies that the additive noise term η_i has been taken care of previously by means of some linear or nonlinear filtering technique, i.e., anisotropic filtering [69], [147], otherwise the model does not apply. However, we recall that in [81] a strong case was made against filtering of the image previously to IIH correction.

2.2.1 GradClassLeg

We call Gradient descent of Classified images corrected by products of Legendre polynomials (GradClassLeg) to our own proposition of a parametric IIH correction and voxel classification algorithm. We assume that the IIH field model is a linear combination of 3D products of Legendre polynomials [26, 157] given by:

$$\beta_i(\mathbf{p}) = \sum_{j=0}^{m} \sum_{k=0}^{m-j} \sum_{l=0}^{m-k-j} p_{jkl} P_j(i_x) P_k(i_y) P_l(i_z);$$ (2.4)

where $i = (i_x, i_y, i_z)$ and $P_k(.)$ is a discretization of the Legendre polynomial of order k that is consistent with the image size in each dimension. The maximum Legendre polynomial order is denoted by m. The expression (2.4) takes into account the symmetries in the composition of the bias, assuming that the discrete volume is a cube. The number of parameters that compose $\mathbf{p} = \{p_{jkl}\}$ is

$n = (m+1) \frac{m+2}{2} \frac{m+3}{3}$. Given an IIH field estimation $\widehat{\beta}_i(\mathbf{p})$, we consider the image correction error relative to the intensity class means as the objective function:

$$e(\mathbf{p},\Gamma) = \sum_{i \in I} \left(\frac{y_i}{\widehat{\beta}_i(\mathbf{p})} - \mu_{x_i} \right)^2 \tag{2.5}$$

where Γ denotes the set of the tissue intensity class means. That is, we compute the restoration error as the squared difference between the predicted intensity associated with the tissue class and the observed intensity after bias correction.

GradClassLeg is a gradient descent algorithm of this error function either on the parameters of the bias field model:

$$\mathbf{p}_{t+1} = \mathbf{p}_t + \alpha_\mathbf{t}^\mathbf{P} \nabla_\mathbf{p} e(\mathbf{p},\Gamma), \tag{2.6}$$

or on the intensity class means:

$$\Gamma_{t+1} = \Gamma_t + \alpha_t^\Gamma \nabla_\mathbf{\Gamma} e(\mathbf{p},\Gamma). \tag{2.7}$$

In these expressions, $\alpha_t^\mathbf{P}$ and α_t^Γ denote the learning rates. Equation (2.6) gives the estimation of the IIH field model parameters starting from a random initial model. Equation (2.7) gives the estimation of the intensity class means, starting from a random initial set of values or from a good guess. The bias field model component of the gradient vector in equation (2.6) is:

$$\nabla_\mathbf{p} e(\mathbf{p},\Gamma) = \left\{ \frac{\partial}{\partial p_{jkl}} e(\mathbf{p},\Gamma) \right\}, \tag{2.8}$$

where:

$$\frac{\partial}{\partial p_{jkl}} e(\mathbf{p},\Gamma) = \sum_{i \in I} \left(\frac{y_i}{\widehat{\beta}_i(\mathbf{p})} - \widehat{\mu}_{\widehat{x}_i} \right) \frac{-y_i P_j(i_x) P_k(i_y) P_l(i_z)}{\widehat{\beta}_i^2(\mathbf{p})}. \tag{2.9}$$

where

$$\widehat{x}_i = \arg\min_\omega \left\{ \frac{y_i}{\widehat{\beta}_i(\mathbf{p})} - \mu_\omega \right\}$$

is the estimation of the classification of each voxel based on the current estimation of the intensity class means. The intensity class means component of the gradient

vector in equation (2.7) is $\nabla_{\mathbf{r}} e\left(\mathbf{p},\Gamma\right) = \left\{\frac{\partial}{\partial \mu_\omega} e\left(\mathbf{p},\Gamma\right)\right\}$ where:

$$\frac{\partial}{\partial \mu_\omega} e\left(\mathbf{p},I\right) = \sum_{i \in I | \widehat{x}_i = \omega} -\frac{1}{2}\left(\frac{y_i}{\widehat{\beta}_i\left(\mathbf{p}\right)} - \widehat{\mu}_\omega\right). \tag{2.10}$$

The estimation of the intensity class means and that of the IIH field model parameters are interdependent processes, they could be performed alternatively, like the ICM and bias estimation performed in the BMAP algorithm, or simultaneously. When we try to estimate both we apply equations (2.6) and (2.7) simultaneously with different learning rates $\alpha_t^{\mathbf{P}}$ and α_t^{Γ} tuned for the different processes.

2.2.2 Discussion of GradClassLeg

In this discussion we will address the following three questions:

1. Why the use of a parametric intensity inhomogeneity model?

2. Does the shape of our energy function, defined in equation (2.5), improve somehow the one proposed by [157]?

3. It is adequate the choice of a gradient descent as the energy minimization algorithm?

Why a parametric IIH algorithm?

Parametric models have the advantages of *compactness* and *abstraction*.

- By *compactness* we mean that the search for the IIH field estimation is performed on the parameter space which is of much smaller dimensionality than the IIH field.

- By *abstraction* we mean that the estimation is not trying to model the small variations of the image, avoiding the kind of overfitting that can happen in non-parametric approaches. *Abstraction* is similar to the regularization of ill-posed problems.

The choice of inhomogeneity model varies widely. In the literature we have found spline interpolations, high order polynomials, products of Legendre polynomials and mixtures of Gaussians. Interpolation approaches such as splines [111, 142]

have the disadvantage of needing some good selections of points or regions in the image, that must be performed by expert human operators. Defining the knot spacing for cubic b-splines needs some degree of user interaction to ensure that the spacing is small enough to model the bias field and greater than the anatomical structures, to avoid that the bias model includes them. Besides, optimization of knot spacing may be required for each medical imaging application. High order polynomial fitting may lead very easily to over-fitting and biased estimations. The products of Legendre polynomials form an orthogonal function basis defined by few parameters. IIH field estimation is performed on the space of the linear combination parameters, which means that the search space is smaller than for other approaches. Other approximations, like the mixture of Gaussian functions need careful selection and have more parameters to estimate (i.e., means, covariance matrices) besides the mixture parameters. There are no definitive criteria and computational results that may allow an objective unbiased selection of the parametric inhomogeneity model.

About the shape of the energy function

The first big difference between ours and other approaches found in the literature that may be noticed is that we avoid to use the logarithm transformation of the image. The main drawback of sticking with the multiplicative model is that low values of the IIH field could give anomalous correction results. However in practice these values are easily masked out to avoid their effect on the gradient estimation. Another effect of the multiplicative inhomogeneity correction is that the magnitude of the bias effectively computed must be small. This condition imposes a restriction on the range of the linear combination parameters to avoid negative or zero IIH field values. In practice we start the random inhomogeneity linear combination parameters in the range $[0.01, 0.001]$. The algorithm maintains these parameter values close to this range.

The second, and more important difference, is that our elementary energy function is the square distance to the class intensity mean

$$e\left(y, \Gamma\right) = \left(y - \mu_{x_i}\right)^2.$$
(2.11)

instead of the diverse versions of the *valley* function proposed in [26, 157]:

$$valley\left(d\right) = \frac{-d^2}{d^2 + 3\sigma^2}.$$
(2.12)

The *valley* function was inspired in the works on robust statistics, i.e., [104] tries to obtain *M*-estimators which are robust against outliers, overcoming the need to specify a scale factor via the realization of a deterministic annealing scheme to follow a Graduated Non-Convexity path in the minimization of a potential function, which would guarantee reaching the global optimal estimation, despite the non-convex nature of the potential function. This annealing was intended to be performed on the potential function. However, in [26, 157] the authors treat the *valley* function of equation (2.12) as a robust estimator whose scale is given by the $sigma$ parameter. Therefore, this parameter is critical to obtain the desired results and its optimal value will be highly dependent on the range of values of the processed signal. On the other hand, the *square* function in (2.11) does not depend on any scale parameter.
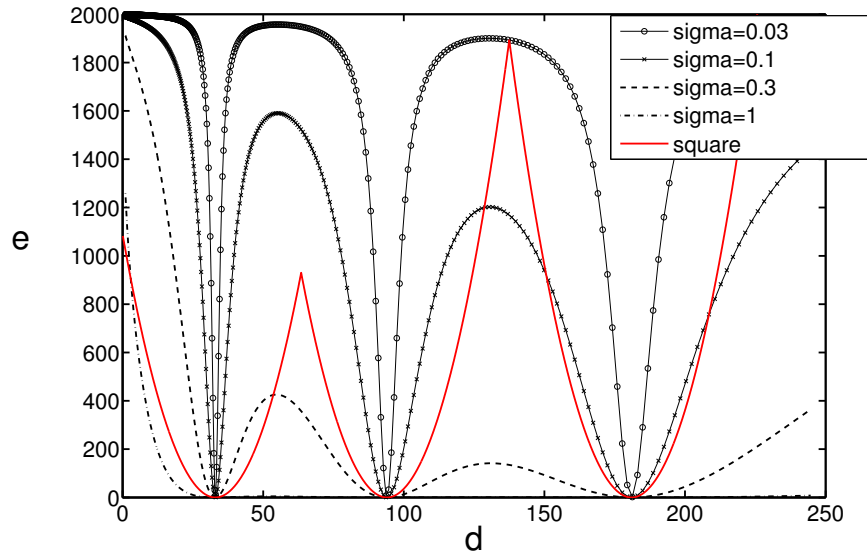
Figure 2.1: The response of the *valley* and *square* functions. Energy plots versus pixel intensity showing the basins of attraction to intensity class means using the square function of equation (2.11) and the valley function of equation (2.12) as building blocks.

One claimed and highly desirable feature of the *valley* function is that it is able to provide, for some values of its σ parameter, separate and well defined attraction basins for the different class means. In figure 2.1 we show both the response of the *valley* and *square* functions when the assumed class mean intensities are $\mu_1 = 33$,

$\mu_1 = 94$ and $\mu_1 = 181$ for input image intensity in the $[0, 255]$ interval. The *valley* function has been scaled multiplying it by 2000 for adequate simultaneous plotting with the *square* function in this figure. Following the procedures described in [26, 157], the *valley* function of equation (2.12) is computed on the logarithm of both the intensity and the intensity class means. The existence of the attraction basins indicates that the energy function gives information about departures from the desired values. Both functions provide this information. Note that for $\sigma = 1$ the *valley* function does not give any information regarding the nominal intensity class means because it is null for most of intensity value range. For $\sigma = 0.3$ the *valley* function is quite similar to the *square* function. The second interesting feature of the the *valley* function is that it saturates to 1. This is natural when we try to filter out outliers. However this non linearity may introduce undesired effects.

(a)   (b)   (c)
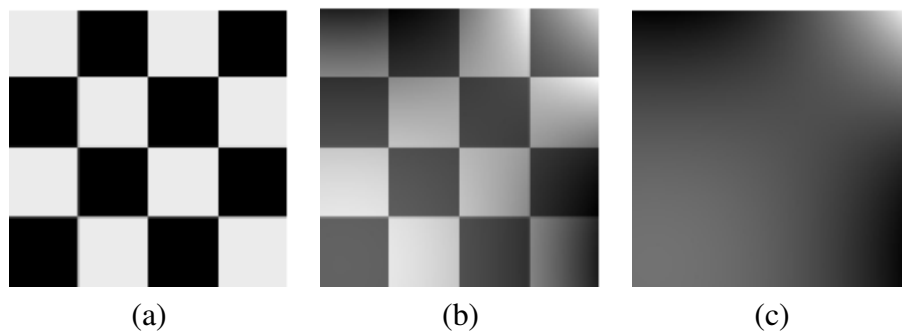
Figure 2.2: Chessboard image. (a) Original. (b) An instance of an IIH corrupted image. (c) The IIH field multiplying (a) to obtain (b).

As an illustrative example of the undesirable effects introduced by the nonlinearities of the *valley* function, we performed the following experiment: starting from a chessboard image like the one in figure 2.2(a), we generated a collection of 10000 IIH corrupted images (as in 2.2(b)) multiplying randomly generated IIH fields based on Legendre polynomials (up to order 2, figure 2.2(c) shows one IIH field) according to equation (2.1). We plot in figure 2.3 the energy function as defined in equation (2.12) [157] with $\sigma = 0.03$ (which, from the examination of figure 2.1, shows a good separation of the attraction basin for the local minima corresponding to the intensity class means) computed for each corrupted image versus its distortion relative to the original chessboard image. Notice that the dispersion of the energy function increases with the image distortion. Figure 2.4 corresponds to a value of $\sigma = 1$. In this figure there is a much smaller dispersion
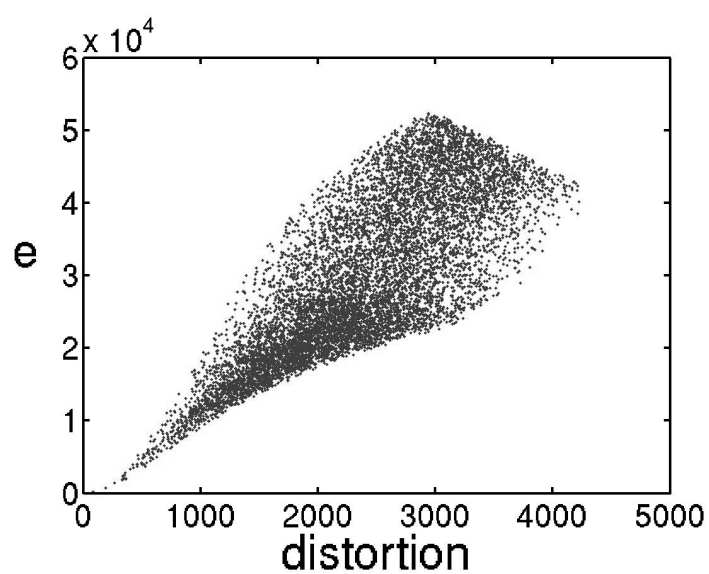
Figure 2.3: Sampling of the energy function based on the *valley* function with $\sigma = 0.03$ of randomly corrupted chessboard images versus their distortion relative to the original chessboard image. IIH fields generated with random Legendre polynomials up to order 2.
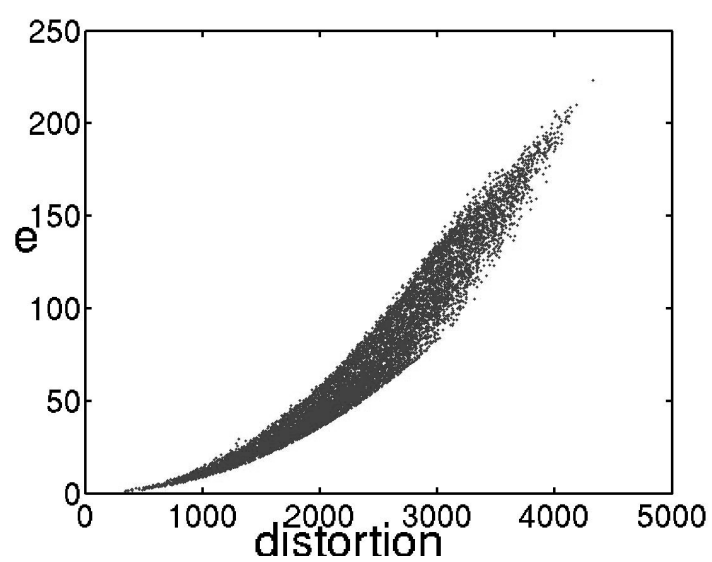
Figure 2.4: Sampling of the energy function based on the *valley* function with $\sigma = 1$ of randomly corrupted chessboard images versus their distortion relative to the original chessboard image. IIH fields generated with random Legendre polynomials up to order 2.
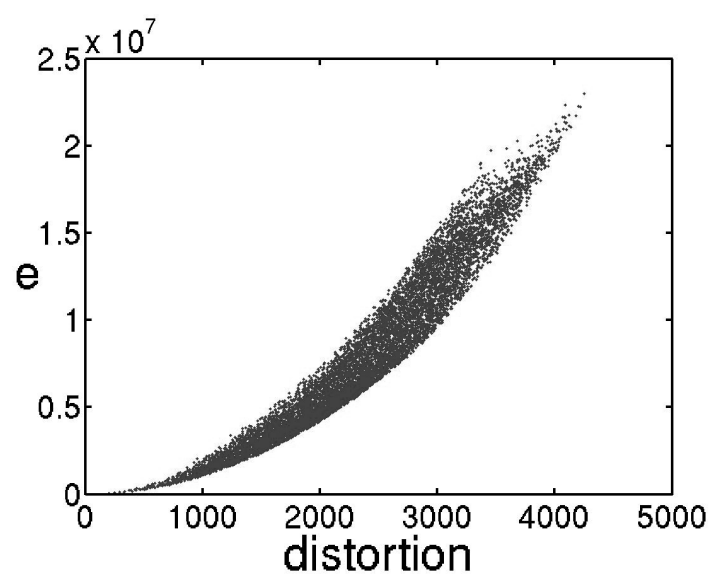
Figure 2.5: Sampling of the energy function based on the *square* function of randomly corrupted chessboard images versus their distortion relative to the original chessboard image. IIH fields generated with random Legendre polynomials up to order 2.

than in figure 2.3. Figure 2.5 corresponds to the plot of the energy function in equation (2.5) versus the corrupted image distortion. We must keep in mind that we search for the minimum distortion (best corrected) image, therefore it is desirable that the energy function used to evaluate the corrupted image is increasingly monotonic with the distortion. The plot of the *square* based energy function is much closer to the plot of a monotonically increasing function than the one based on the *valley* function. This is due to the saturation of the *valley* function. The dispersion shown in figure 2.3 means that there are images with widely varying distortion values that have the same energy value, so that minimization algorithms using this energy as the objective function may be far from searching for the minimum distortion. The picture gets worse for higher order polynomials, increased number of classes, etc. Moreover, the magnitude of this effect varies widely with the value of the σ parameter, as can be appreciated comparing figure 2.3 and figure 2.4. The equivalence between the *valley* and the *square* functions for specific values of σ can be appreciated by the similitude between figures 2.5 and 2.4.

About the gradient descent

It seems from the simple example explored in [157] that the minimization problem is highly non linear and the search surface has many local minima. To analyze the convexity of the energy function proposed in equation (2.5) we can examine the expression of the second derivatives:

$$\frac{\partial^2 e\left(\mathbf{p},\Gamma\right)}{\partial^2 p_{jkl}} = \sum_{i\in I}\widehat{\mu}_{\widehat{x}_i}\frac{y_i\left(P_j\left(i_x\right)P_k\left(i_y\right)P_l\left(i_z\right)\right)^2}{\widehat{\beta}_i^3\left(\mathbf{p}\right)}. \tag{2.13}$$

$$\frac{\partial^2 e\left(\mathbf{p},\Gamma\right)}{\partial p_{jkl}\partial p_{mno}} = \sum_{i\in I}\frac{\widehat{\mu}_{\widehat{x}_i}y_iP_j\left(i_x\right)P_k\left(i_y\right)P_l\left(i_z\right)P_m\left(i_x\right)P_n\left(i_y\right)P_o\left(i_z\right)}{\widehat{\beta}_i^3\left(\mathbf{p}\right)} \tag{2.14}$$

A sufficient condition for the Hessian matrix to be positive semidefinite and, therefore, for the energy function to be convex, is that all the second derivatives are positive anywhere. From these equations, we can deduce that sufficient conditions for convexity of the energy function are:

1. That the IIH field is non negative for every image element, and

2. That the elementary polynomials $P_j\left(.\right)$ are also non negative.

The first condition correspond to a realistic physical restriction if we maintain the multiplicative inhomogeneity model. To enforce the second condition implies displacing the polynomial value range from $[-1, 1]$ to $[0, 1]$. This causes the loss of orthogonality of the modified basis. As this property plays no role in the proposed algorithm, we can perform this translation without fear of introducing unsuspected problems. Under these conditions the gradient descent algorithm proposed is a global minimization algorithm and no further sophisticated random search algorithms are needed to reach the global minimum.

2.2.3 Adaptive field rule (AFR)

Assuming the image formation model shown in equation (2.1), the image segmentation problem, subjected to the existence of IIH multiplicative noise, can be formulated as the minimization of the following energy function:

$$E(\mathbf{y};\Gamma,\beta) = \sum_i \left(\frac{y_i}{\beta_i} - \mu_{c(i)} \right)^2,$$ (2.15)

where

$$c\left(i\right) = \arg\min_k \left\{ \left\| \mu_k - \frac{y_i}{\beta_i} \right\| \right\}.$$

That is, to solve the segmentation problem we need to find

$$\Gamma^*, \beta^* = \arg\min_{\Gamma,\beta} E(\mathbf{y};\Gamma,\beta).$$

When the intensity class means $\Gamma = \{\mu_{\omega_1}, ..., \mu_{\omega_c}\}$ are known, the problem is reduced to that of IIH estimation:

$$\beta^* = \arg\min_\beta E(\mathbf{y};\Gamma,\beta).$$

That is, we try to minimize the quantization error over the IIH corrected image. This is an ill-posed problem because the number of parameters to estimate (the bias values at each voxel) is the same as the number of data samples (the values of the voxel intensities).

The logarithm is a monotonic transformation, therefore the segmentation problem can be stated as the following minimization problem:

$$\mathbf{B}^* = \arg\min_{\mathbf{B}} \sum_i \left(Y_i - B_i - M_{c(i)} \right)^2 = \arg\min_{\mathbf{B}} E\left(\mathbf{Y}; \mathbf{B}, \mathbf{M}\right)$$ (2.16)

where $Y_i = \log y_i$, $B_i = \log \beta_i$ and $M_{c(i)} = \log \mu_{c(i)}$. This logarithmic transformation, which allows us to get rid of the multiplicative effect, is quite common in the MRI segmentation and IIH correction literature [127, 149, 168]. We consider a gradient descent rule for the minimization of this error function. We obtain the following adaptive estimation rule:

$$\triangle B_i = -2\alpha \left(Y_i - B_i - M_{c(i)} \right) \tag{2.17}$$

where $0 < \alpha < 1$ as usual. To take into account that IIH fields must be smooth, we propose an energy function formulation that involves the smoothing (averaging) of the errors in the neighborhood of the voxel:

$$E_\sigma \left(\mathbf{Y}; \mathbf{B}, \mathbf{M} \right) = \sum_i \sum_j h_\sigma \left(i - j \right) \left(Y_j - B_j - M_{c(j)} \right)^2 \tag{2.18}$$

The energy function E_σ is an extended distortion function like the energy functions proposed for the derivation of topological preservation quantization and visualization approaches [84, 51, 161, 103] inspired by the SOM [96] and the Neural Gas. The $h_\sigma \left(i - j \right)$ is a neighboring function defined in the space of the voxel index sites. Usually it is a Gaussian shaped windowing function.

The formulation of a gradient descent rule for the minimization of energy function E_σ gives the following the Adaptive Field Rule (AFR):

$$\triangle B_i = -2\alpha \sum_j h_\sigma \left(i - j \right) \left(Y_j - B_j - M_{c(j)} \right). \tag{2.19}$$

The AFR rule of equation (2.19) can be applied either as stochastic gradient descent or as batch rule. We have tested both in [66], however here we report results applying AFR as a batch rule. The main reasons for this choice are, first, that the batch version has faster convergence both in computer time and in number of iterations, and, second, that the batch version has lower result variance.

2.2.4 Discussion of AFR

Similarities and differences with SOM

Although the AFR as defined in equation (2.19) was inspired by the SOM's weight adaptive rule, there are some strong differences. The first one is paradigmatic: the SOM is a quantization algorithm that achieves a clustering of high-dimensional

data and a nonlinear analog to the PCA because of its topology-preserving property. Thus, the number of network nodes is at least several orders of magnitude smaller than the data sample. With AFR we try to perform a non-parametric estimation of the IIH field, which can be viewed as an image of the same size as the given MRI image which constitutes our data sample. The updating process is a one to one process. Each voxel of the MRI image is visited (in sequence or randomly) and its corresponding IIH field element is updated.

The role of the neighboring function in AFR as introduced in equation (2.19) works in the opposite direction to the one in the standard SOM. In the standard SOM the winning codevector determines the neighboring codevectors being updated. In equation (2.19) the updating of the bias estimated value at a voxel is influenced by the error at neighboring voxels.

Finally, let us consider the topographic preservation property of SOM ([57, 166]). According to it, neighboring data points in input space are mapped into network nodes with neighboring indices. The definition of AFR in equation (2.19) pursues that neighboring voxels have estimated IIH field neighboring values to ensure the smoothness of the bias field. Voxel site indices are defined in a 2D or 3D space and bias values are scalar values, much unlike to the SOM, where dimension reduction is produced because the codevector indices are defined in a space of lower dimension than that of the input vectors.

Discussion of convergence properties

There are several causes for the fact that the observed intensity is different from the corresponding intensity class mean $Y_j \neq M_{c(j)}$. Among them, the most important is the partial volume phenomena, due to the fact that image voxels correspond to a volume of matter that contains a mixture of biological tissues. Variations of this mixture from voxel to voxel produce variations in the signal departing from the nominal value, although the tissue class (i.e., CSF, GM or WM) may be the same. These and other uncorrelated signal variations can be accounted for by the additive term in the model of equation (2.1). There are other variations due to smooth spatial processes, such as the spatial inhomogeneity of the RF signal that excites the protons in the selected volume to produce the MRI signal [46]. The estimated IIH field $\hat{B}$ must be smooth to be a model that accounts for such physical phenomena [1]. This smoothness constraint can be expressed in several ways

---

[1]The assumption of negligible additive noise in log transformation of equation (2.3) forces some confusion additive and multiplicative effects. In fact, in some works, i.e., [157], it is unclear

(expressed in terms of the IIH field before/after the log transform):

1. The variation of the bias must be bounded. That can be expressed by either expression

$$\max_{j \in N(i)} \|\beta_i - \beta_j\| < \varepsilon,$$

or

$$\sum_{j \in N(i)} \|\beta_i - \beta_j\| < \varepsilon.$$

2. Closer voxels in the image domain must have more similar bias values than those farther appart:

$$\forall i, j, k; |i - j| < |i - k| \Rightarrow \|\beta_i - \beta_j\| < \|\beta_i - \beta_k\|. \tag{2.20}$$

This condition means that there is some kind of topological preservation between voxel site indices and bias values.

The application of equation (2.19) leads to stationary states described by the following equation:

$$\hat{B}_i = \sum_j h_\sigma (i - j) \left(Y_j - M_{c(j)} \right), \tag{2.21}$$

that is, the estimation corresponds to the weighted average of the neighboring voxel residuals. This asymptotic result is independent of the shifts in $c(j)$ that may be due to the variation in values of the bias $\hat{B}_i$ during the estimation process. The simplistic appearance of equation (2.21) hides a recurrent relation between the bias estimations given by the class assignment

$$c(i) = \arg \min_k \left\{ \left\| M_k - \left(Y_i - \hat{B}_i \right) \right\| \right\} \tag{2.22}$$

If the neighboring function is an average mask $h_\sigma(k) = 1/\sigma^d$ for $|k| \leq \sigma$, where d is the dimension of the image support (2 or 3), the $\hat{B}_i$ would correspond to the average residual over a box or cube surrounding the voxel. Maintaining this definition of the neighboring function, the justification of the satisfaction of the smoothness constraints on the bias estimate after convergence of the updating rule come from:

$$\left| \hat{B}_i - \hat{B}_j \right| = \sum_{k \in N(i) \triangle N(j)} \left| Y_k - M_{c(k)} \right|, \tag{2.23}$$

---

wether the additive correction/estimation was done on the original image or its log transform.

where $A \triangle B$ stands for the symmetric difference between sets A and B, $N(i)$ is the set of voxel neighboring sites. Then, the magnitude of the difference will always be bounded, but we can not set a bound because this difference will depend on the input image. The topological preservation condition of equation (2.20) is also satisfied because

$$\forall i, j, k; |i - j| < |i - k| \Rightarrow N(i) \triangle N(j) \subset N(i) \triangle N(k).$$

When the neighboring function is a Gaussian or any other arbitrary normalized spatial function with restricted domain of radius $N_\sigma{}^2$, the equation (2.23) becomes

$$|B_i - B_j| = \sum_{k \in N_\sigma(i) \triangle N_\sigma(j)} |Y_k - M_{c(k)}| |h_\sigma(i - k) - h_\sigma(j - k)|.$$

In this case, the decreasing nature of the neighboring function introduces an attenuation of the bias differences making the transition between bias values even more smoother.

The conventional application of the SOM and other neighborhood based competitive algorithms usually involves the progressive shrinking of the neighborhood radius, until it becomes null and the adaptive rule only applies to the actual winning or selected node. The conventional interpretation of this process is that the SOM and similar algorithms become the Simple Competitive Learning or Vector Quantization to allow for refinement of the final codevectors. The big initial neighborhoods ensure global convergence or robustness against initial conditions [44, 74, 78]. In the AFR algorithm, the final neighborhood determines the closeness to the assignment of the bias values as the residual intensities, which is the trivial minima of the error function in equation (2.15). The initial neighborhood size sets the background smoothness constraint. The shrinking neighborhoods allow refining the IHH field estimation adapting to local conditions. The shrinking schedule is of importance, because a long number of iterations with the null neighborhood will remove the smoothness obtained in the initial steps. This is the reason of the selection of the value for the shrinking speed parameter F defined in equation (2.29) of section 2.4.

On the estimation of the intensity class means

Sometimes the assumption of the knowledge of the intensity class means M_k can not be made. Then some means are needed to estimate them from the data. In

---

[2]For Gaussian neighborhoods we apply the rule of thumb $N_\sigma = 9\sigma$ to ensure appropriate sampling of the Gaussian function.

this section we will discuss some approaches that can be combined with the AFR IIH estimation algorithm. These approaches are presented here in sequence of increasing requirements of *a priori* information about the data. We start with the basic unsupervised approach, ending with the supervised learning algorithm. One of the straightforward approaches is the realization of a k-means step after each step of IIH field estimation. This approach has already been tested in section 2.5.1.

This estimation can also be performed as a gradient descent on the energy function of equation (2.18) relative to the intensity class means:

$$\frac{\partial E_\sigma}{\partial M_k} = -2 \sum_i \sum_{c(j)=k} h_\sigma \left(i - j\right) \left(Y_j - B_j - M_k\right).$$

This gradient is well approximated by the following expression, which minimizes the average square error of the image quantization, given a fixed bias field:

$$\frac{\partial E_\sigma}{\partial M_k} \simeq -2 \frac{1}{\left|\{j \,|c\left(j\right) = k\}\right|} \sum_{c(j)=k} \left(Y_j - B_j - M_k\right). \qquad (2.24)$$

The process of estimating the intensity class means following the gradient in equation (2.24) is an unsupervised approach, without *a priori* information about the true class of the voxel's tissue. We will denote AFR-U this approach in the experimental results of sections 2.5.2 and 2.5.3.

Sometimes we may have some information about the spatial distribution of the tissues, this usually takes the form of an atlas[3]. A crisp atlas provides us a tissue class for each voxel. The gradient of equation (2.24) can be rewritten to profit from this information:

$$\frac{\partial E_\sigma}{\partial M_k} \simeq -2 \frac{1}{\left|\{j \,|A\left(j\right) = k\}\right|} \sum_{\{j|A(j)=k\}} \left(Y_j - B_j - M_k\right), \qquad (2.25)$$

where $A\left(j\right)$ denotes the class assigned to voxel site j in the crisp atlas. We denote AFR-A this approach in the experimental section 2.5.3. The crisp atlas is computed following a majority voting schema over a set of images representative of some population of interest. A more precise representation is a probabilistic map, where we have the estimated frequency of each tissue for each voxel computed

---

[3]In the literature, some authors define templates as average images and atlases as average images after warping into a known anatomical atlas, such as the Talairach atlas. Here we use the term to denote *a priori* information about the spatial distribution of tissue classes.

from the image database. The gradient expression for the intensity class means can then be rewritten as follows:

$$\frac{\partial E_\sigma}{\partial M_k} \simeq -2 \frac{1}{\sum_j AP(j,k)} \sum_j AP(j,k)(Y_j - B_j - M_k),$$ (2.26)

where $AP(j,k)$ denotes the estimated frequency of tissue class k in voxel site j. We denote AFR-PA this approach in the experimental section 2.5.3. Both atlas based algorithms are partially supervised because they use some *a priori* information about the expected distribution of the tissue classes in the image, computed from the (manual) segmentation of the images. The fully supervised approach would use the manual segmentation of the image being processed as the ground truth, so the gradient expression can be written as follows:

$$\frac{\partial E_\sigma}{\partial M_k} \simeq -2 \frac{1}{|\{j\,|GT(j) = k\}|} \sum_{\{j|GT(j)=k\}} (Y_j - B_j - M_k),$$ (2.27)

where $GT(j)$ denotes the ground truth class at voxel j given by the manual segmentation of the image being processed. We will denote AFR-S this approach in the experiments reported in section 2.5.3.

Rationale for AFR

Non parametric IIH correction methods are able to model local features, such as partial volumes or very local inhomogeneity effects. Parametric models are well suited to model smooth global IIH fields, but modeling of local effects would require extremely high order models and very sensitive fitting algorithms. The AFR is a non parametric algorithm whose degree of locality is controlled by the neighboring function evolution parameters. It could be possible, therefore, to tune it to solve a diversity of problems. Results reported in this chapter show that it can be easily tuned to perform comparable to state of the art algorithms on conventional benchmark data sets.

From the computational point of view, AFR is fast and easily tuned, relative to other non parametric algorithms. Approaches based on Bayesian Maximum a Posteriori estimation [127] involve the computation of relaxation processes, which, even in the fastest ICM case, are time consuming. The early Bayesian approaches [168] also needed to perform some linear filtering processes. Algorithms based on fuzzy clustering [18, 148] need to compute membership functions and solve large

sets of equations to obtain them. Algorithms based on the intensity probability distribution sharpening [149] need to perform a sequence of distribution deconvolutions. All of these processes are more computationally demanding than AFR. We have found that a small number of iterations, 20 to 30, are enough to obtain good results for most volumes treated.

From the modeling point of view, the Bayesian approaches need to build a probabilistic model, often a Markov Random Field, whose parameters need to be estimated somehow. Otherwise AFR provides a dynamic process along the sensitive parameter values so that the response is less dependent on a critical parameter choice. The process works like an annealing process (it can also be interpreted as a graduated non convex minimization or a continuation problem). This allows fitting it to solve global and local problems. To obtain this flexibility, the Bayesian methods need to work on the *a priori* model of the data. Fuzzy approaches have introduced averaging processes to obtain similar results, the width of the averaging window is set heuristically.

The assumption of known intensity class means is common to several approaches [157, 168]. In the fuzzy c-means and Bayesian approaches, intensity class means are computed all along, as the corrected image evolves. We show in this chapter that several approaches can be combined with AFR to estimate the intensity class means with great efficiency. The simplest ones are the unsupervised on-line and k-means processes. The construction of an atlas from the available manual segmentations allows to manage spatial distribution information, like in the template based Bayesian approaches [111]. However, our experience described below shows that the use of templates or atlas must be taken with caution. Supervised training can also be combined with AFR aiming to obtain more accurate segmentations.

2.3   Computational experimentation framework

The validation of the IIH correction algorithms has been done on volumes obtained from two Internet sites:

- The BrainWeb web site [4] at the McConnell Brain Imaging Center of the Montreal Neurological Institute, McGill University, that provides simulated brain volumes [98], The advantage of working with the simulated volume is that it is possible to compute the classification accuracy relative to the ground truth classes effectively defined in the generation model. References [40, 41, 97, 98] develop the foundations of the simulation provided

there. The volumes are generated from a normal anatomic model using a T1 sequence, corrupted with intensity inhomogeneities of magnitude 20% and 40% of the original clean image, and several levels of additive noise. See Appendix C for more details.

- The Internet Brain Segmentation Repository [8]: The 20 normal MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital. This dataset has been extensively used for validation of segmentation approaches in the literature [21, 65, 111, 127, 148, 175]. The images correspond to T1-weighted MRI brain scans of 20 normal subjects[4]. The images were obtained with two different imaging systems[5]: ten FLASH scans of four males and six females, ten 3D-CAPRY scans on six males and four females. The differences in the imaging systems drive some authors to consider them separately or to omit some of the scans when reporting results [21, 127]. The IBSR V2.0 collection of brain volumes, the 1.5mm data distribution, is more recent than the previous one, and with less reference results. Again, the MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital. We have used it for the AFR algorithm validation. See Appendix C for more details.

To give a quantitative evaluation of GradClassLeg we have computed the Tanimoto coefficient for the CSF, GM and WM classes as introduced in [111, 148]. Using our notation the Tanimoto coefficient can be expressed as

$$T(\omega) = \frac{|\{x_i = \omega \wedge \widehat{x}_i = \omega; i \in I\}|}{|\{x_i = \omega \vee \widehat{x}_i = \omega; i \in I\}|}, \tag{2.28}$$

---

[4]We discarded one of the scans because the manually segmented volume did not seem to match with the T1 dataset applying the given offset.

[5]MRI Image Acquisition description: The coronal three-dimensional T1-weighted spoiled gradient echo MRI scans were performed on two different imaging systems. Ten FLASH scans on four males and six females were performed on a 1.5 tesla Siemens Magnetom MR System (Iselin, NJ) with the following parameters: TR = 40 msec, TE = 8 msec, flip angle = 50 degrees, field of view = 30 cm, slice thickness = contiguous 3.1 mm, matrix = 256x256, and averages = 1. Ten 3D-CAPRY scans on six males and four females were performed on a 1.5 tesla General Electric Signa MR System (Milwaukee, WI), with the following parameters: TR = 50 msec, TE = 9 msec, flip angle = 50 degrees, field of view = 24 cm, slice thickness = contiguous 3.0mm, matrix = 256x256, and averages = 1.

where x_i is the voxel's ground truth and $\widehat{x}_i$ is the classification estimation performed by the algorithm. This definition is slightly different from the classical one found in the literature [49] which is the ratio of cardinalities of the symmetrical difference and the union of the sets. In equation (2.28) perfect overlapping gives a value of 1 while disjoint sets give a value of 0. Perfect overlapping means the perfect agreement between the manual and automatic segmentation. It is the same statistic as the Overlap Metric reported in the IBSR site. The Tanimoto coefficient is a more precise measure of the algorithm accuracy than the ratio of success or the misclassification ratio (MCR) given by some authors, because it takes into account the error due to the false positives as well as the false negatives. In all the experimental works, we computed the classification of the CSF, GM and WM, and reported the Tanimoto coefficients for these classes. When visualizing the voxel classification results, the darkest gray correspond to CSF tissues and the whitest voxels correspond to WM tissues. Intermediate gray values correspond to the GM tissues.

2.4   Computational results for GradClassLeg

We present in this section comparative computational results of the 3D implementations of the GradClassLeg algorithm described in section 2.2.1 relative to state of the art algorithms. In this section we show empirically that the gradient descent algorithm is able to obtain good estimations of the IIH field, and that it is comparable to state of the art algorithms. The experimental data is composed of two sets of images: one set are simulated brain scans from the BrainWeb, and the second set are real volumes from the IBSR (20 normal MR brain data sets and their manual segmentations). The treatment of the voxels corresponding to air in the scanned volumes was different in the two experiments, and depending on the algorithm. On the BrainWeb images, for the Wells and BMAP algorithms air was considered as a class of its own. For the GradClassLeg and BFCM air was detected setting a threshold and voxels below this threshold were ignored in all subsequent computations. On the IBSR images the air was masked out using the manual labeling of the voxels and was ignored. We use the simulated volumes to compare the restoration power of the algorithms, and the IBSR data to compare their segmentation power.

2.4.1 Results on simulated brain MRI volumes

We have downloaded a simulated clean normal brain volume, without additive or intensity inhomogeneity noise, of a healthy head model and we have corrupted it with randomly generated IIH fields based on Legendre polynomials to obtain the test images. We have generated a hundred corrupted volumes. Our bias fields are stronger than the ones introduced in [4] so that the relative differences between algorithms are enhanced. Working with the simulated volume it is possible to evaluate the algorithms on the basis of the correlation between the original, clean, volume and the corrected volume after applying the estimated IIH field.

The algorithm presented in section 2.2.1 and the algorithms described in Appendix A have been used to estimate the IIH field. The corrupted image is then corrected with the estimated IIH field. We have assumed as the true values of the intensity class means the averages of the simulated image regions corresponding to each kind of tissue according to the crisp head model provided in [4]. We consider WM, GM, and CSF, as the tissues of interest. The accuracy of the algorithms is measured computing the correlation of the clean volume simulation and the restored volumes. The operational parameter values applied to each algorithm are the ones recommended or used in the reference papers. They are as follows:

- GradClassLeg: $\alpha^{\mathbf{p}} = 0.01$, 25 iteration steps, maximum Legendre polynomial order $m = 5$.

- BMAP: 3D voxel neighborhood radius $r = 2$, MRF potential parameters $\alpha_1 = 0.1, \alpha_2 = 0.01$, 3D isotropic Gaussian smoothing filter standard deviation $\sigma = 20$.

- Wells algorithm: 3D isotropic Gaussian smoothing filter standard deviation $\sigma = 25$.

- BFCM: 3D voxel neighborhood radius $r = 1$.

Figure 2.6 shows one slice of the clean simulated brain volume (fig. 2.6(a)), together with the same slice extracted from one of the IIH corrupted volumes (fig. 2.6(b)). Figure 2.7 shows the results of the IIH correction with the bias field estimations performed by the algorithms described in the previous section, computed on the corrupted volume whose slice is shown in figure 2.6(b). Figures 2.7(a-d) show the corresponding slices of the corrected volumes, after correcting with the IIH fields estimated by GradClassLeg, Wells, BMAP and BFCM algorithms.

(a) (b) (c)
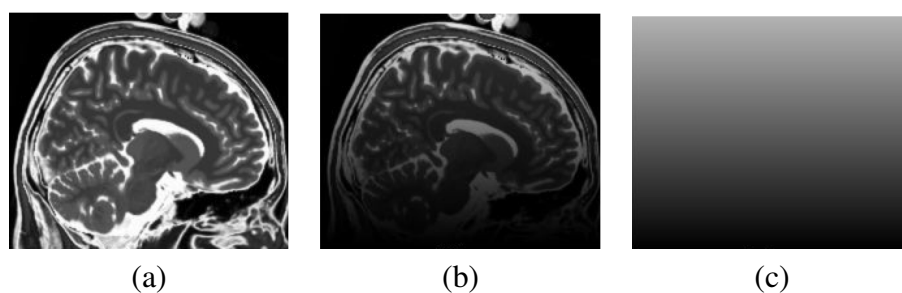
Figure 2.6: A simulated brain MRI slice. (a) The clean slice, (b) The same slice of the volume corrupted with a simulated IIH field from random linear combinations of Legendre polynomial products.

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

(e) (f) (g) (h)
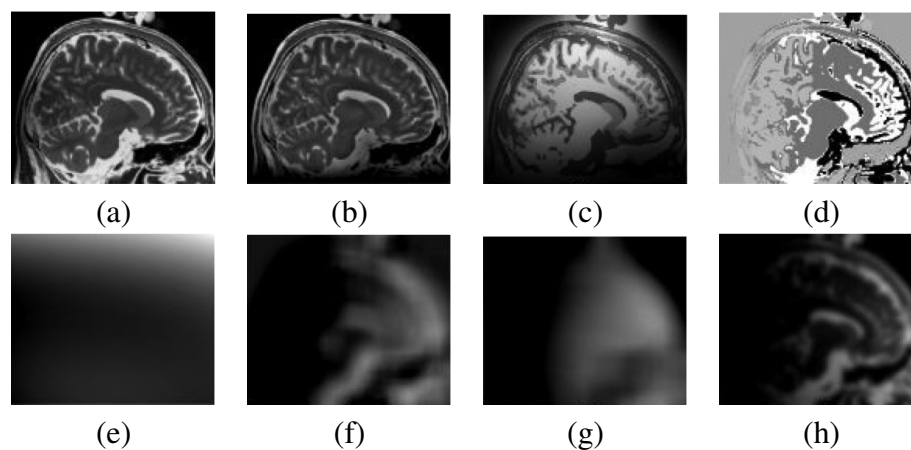
Figure 2.7: First row shows IIH correction results for the volume represented by the slice image in figure 2.6. Second row shows the estimated bias fields. The columns correspond, from left to right, to GradClassLeg, Wells, BMAP and BFCM.
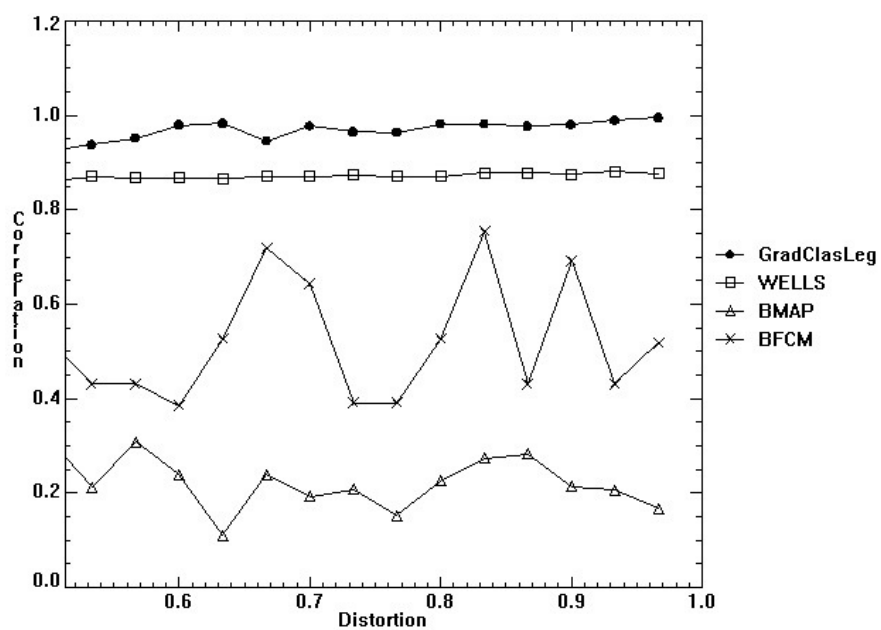
Figure 2.8: Correlation of IIH corrected volumes with the clean volume for a collection of volumes of increasing intensity inhomogeneity.

Figures 2.7(e-h) show the corresponding slice of the estimated multiplicative IIH fields. White values correspond to IIH field values close to 1.

Upon inspection of figures 2.7(a-d) it can be appreciated that the best visual correction corresponds to the Wells algorithm and GradClassLeg. Visual differences between both algorithms are small. The restoration results of the BMAP are very poor. The corrected image looks like a poor quantization of the original image, with some spurious boundary effects that create a kind of aura around the head. We believe that the reasons for this bad behavior is that the parameter setting discussed in the original paper is not adequate for our test images. The BFCM corrected image appears to be an image of differences around the local image intensity mean. This may be due to the fact that the estimated bias is computed like an image of residuals according to equation (A.17). It seems that the smoothing implicit in the neighboring term is too weak for our test images.

The plots in figure 2.8 contain the correlation of the IIH corrected images with the original image, for a collection of IIH corrupted images of increasing distortion relative to the original image. In the figure, the x-axis corresponds to the correlation between the original image and the corrupted image. The y-axis corresponds to the correlation between the original image and the corrected image. It can be appreciated that GradClassLeg systematically improves over the other algorithms. These results confirm that the gradient descent approach assumed by GradClassLeg is robust enough to estimate strong IHH fields.

2.4.2 Sensitivity analysis of GradClassLeg

The aim of the experiments reported in this section is to show the robustness of the GradClassLeg algorithm against bad initializations of the intensity class means estimations.

The experimental data is composed of two simulated brain MRI volumes [98] obtained from the BrainWeb. Using the class information given in the site we have masked out the pixels not belonging to the three classes of interest: WM, GM, and CSF. We have also downsampled the volume to obtain a smaller volume to allow for extensive experiments. The GradClassLeg parameters are set to $\alpha^{\mathbf{P}} = 0.01, \alpha^{\Gamma} = 0.1$, and the maximum number of iterations allowed is 100. Initial IIH field parameters estimation $\hat{\mathbf{p}}$ is set to zero. The initial intensity class means $\hat{\Gamma}$ are generated with uniform probability in the interval [0,100] and ordered in ascending order to preserve the meaning of the classes for visualization and computation of validation indices purposes. The natural ascending order of intensities is CSF, GM and WM. If this order corresponds to ascending number

of class, the visualization will not need further labelling neither for the human inspection nor for the computation of the Tanimoto index.
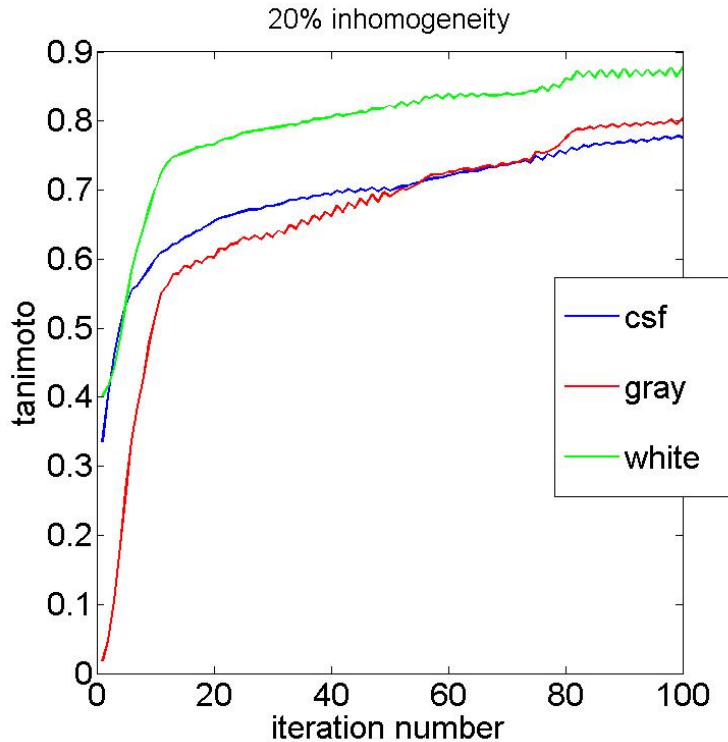
Figure 2.9: Average evolution of the Tanimoto coefficients for CSF, GM and WM obtained over a simulated 20% IIH inhomogeneity brain volume.

The results show in the figures 2.9, 2.10, 2.12 and 2.13 are the evolution of the Tanimoto coefficients values as the algorithm proceeds. The coefficients are computed at each iteration step. Figures 2.9 and 2.12 present the averages of 50 experiments with different random initial means applying the equations (2.6) and (2.7) once at each iteration to the volumes with 20% and 40% inhomogeneities, respectively. It can be appreciated that the algorithm recovers from very bad initial conditions up to acceptable results if we take into account that using the correct intensity class means, computed using the class information from the model, the Tanimoto coefficients obtained are between 0.92 and 0.94. Upon observation, one working hypothesis is that random initial means that lie very close lead to bad results. In figure 2.11 we present the plot of the the Tanimoto coefficients obtained
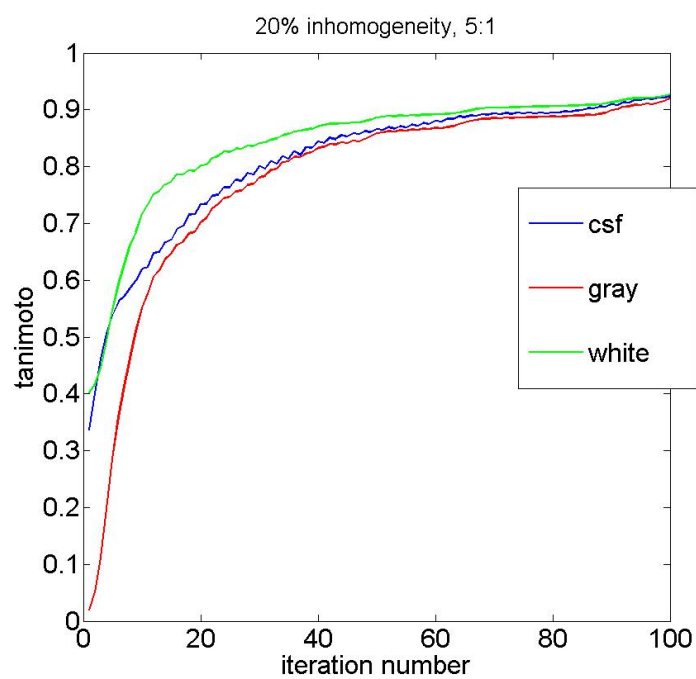
Figure 2.10: Average Tanimoto coefficients for CSF, GM and WM obtained over a simulated 20% IIH inhomogeneity brain volume. Gradient frequency 5:1 (see text).
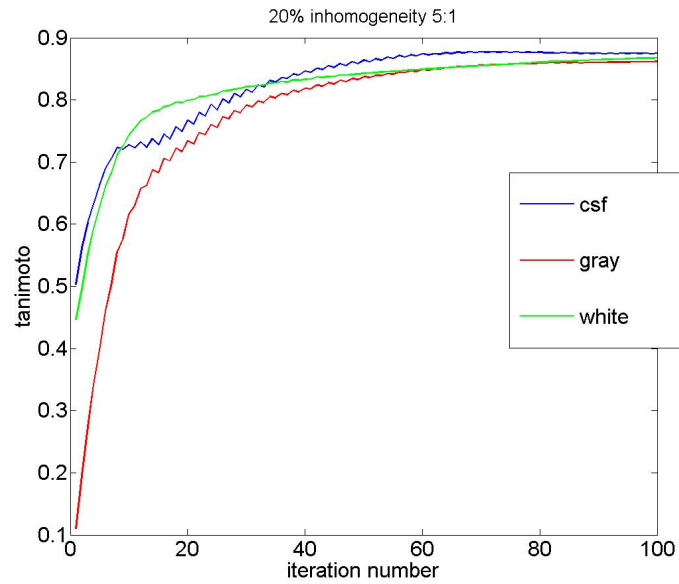
Figure 2.11: Values of Tanimoto coefficients at iteration 100 versus minimum distance between initial class means. 250 instances of the execution of the algoritm. Simulated 20% IIH inhomogeneity.
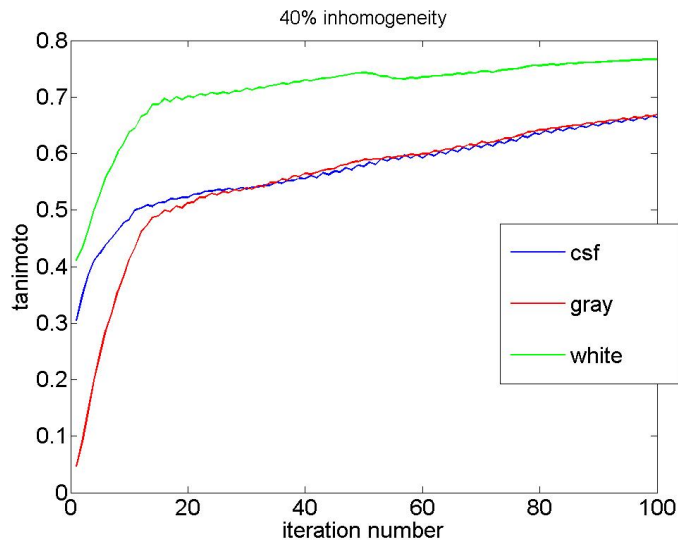
Figure 2.12: Average Tanimoto coefficients for CSF, GM and WM obtained over a simulated 40% IIH inhomogeneity brain volume.
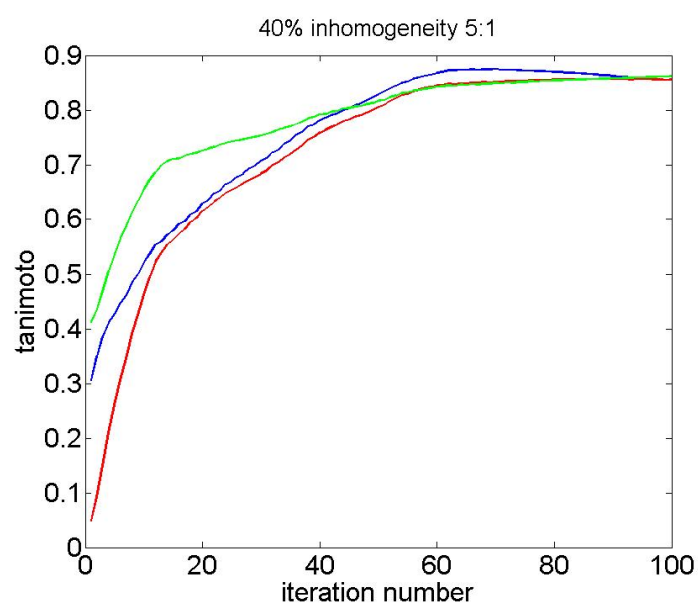
Figure 2.13: Average Tanimoto coefficients for CSF, GM and WM obtained over a simulated 40% IIH inhomogeneity brain volume. Gradient computation relative frequency 5:1 (see text).

at the end of each of 250 replicas against the minimum distance between the initial means. It seems to confirm this intuition because low values of the intraclass distance produce frequent bad results. Finally, in figures 2.10 and 2.13 we try to test the effect of a more frequent mean estimation. The equation (2.7) was applied five times for each computation of the IIH rule (2.6). It can be appreciated that the results improve over the figures 2.9 and 2.12.

For a further quantitative evaluation of the performance of the algorithm, we compute the misclassification rate (MCR) for the brain volumes corresponding to 3% noise and 0%, 20% and 40% inhomogeneity, performing comparisons with the results reported in [107] for a variety of algorithms. These quantitative results are presented in Table 2.1. FCM denotes the conventional fuzzy c-means algorithm. FM -AFCM and TM-AFCM denote the full multigrid adaptive FCM algorithm and the truncated multigrid adaptive FCM algorithm, respectively [124]. EM1 and EM2 denote the unsupervised EM algorithm for finite Gaussian mixture models, where EM1 refers to the standard model and EM2 refers to the model where variances and mixture coefficients of the Gaussian components are assumed equal [106]. AMRF denotes the adaptive Markov random field algorithm [174], [73]. MNI-FCM denotes the method where the N3 inhomogeneity correction technique [149] from MNI is applied first, followed by FCM segmentation. AS-FCM denotes the adaptive spatial FCM algorithm [107]. From Table 2.1, we can see that MCR increases with growing IIH level for all the methods. Our algorithm has better performance than other pure unsupervised algoritms and it is competitive with state-of-the-art algorithms despite its bad initialization and its unsupervised nature. For 0% IIH level, the proposed algorithm has significantly better performance than supervised and unsupervised methods. When studying the results in Table 2.1 it must be taken into account that most works reported there work on a 2D slice basis while our algorithm is a fully 3D algorithm. Besides, the results seem to refer to a single execution of the algorithm with well chosen initial conditions, whereas the result that we report in Table 2.1 comes from a random initialization.

Finally we provide some visual results that demonstrate our results visually. Figure 2.14(a) shows an slice ($z = 90$) of the 3% noise and 20% IIH volume. Figure 2.14(b) shows the corresponding class labeling in the true model. The visual evaluation of the algorithm is given by the estimated bias field shown in figure 2.15(a), and the classification obtained by our algorithm shown in figure 2.15(b). It can be appreciated that the algorithm provides state of the art results.

(a) (b)
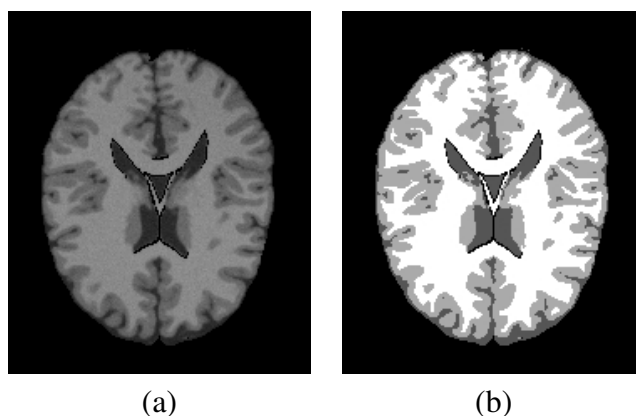
Figure 2.14: (a) Original coronal slice (z=90) of the 3% noise and 20% IIH volume, (b) class labelling in the simulation model of the same slice.

(a) (b)
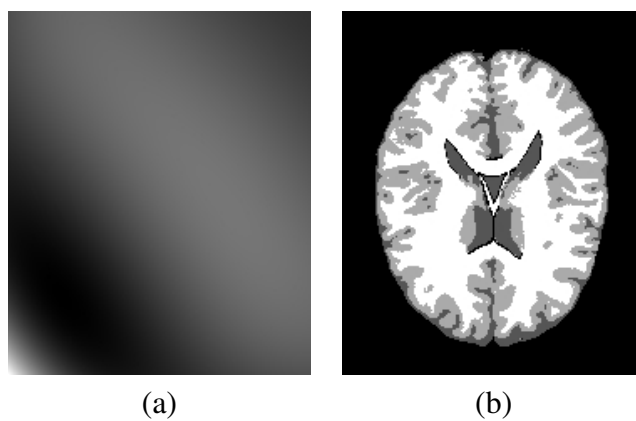
Figure 2.15: Visual evaluation of the GradClassLeg: (a) Estimated bias slice (z=90), (b) Estimated classification of the voxels for the slice of figure 2.14

Method	0% IIH	20% IIH	40% IIH
FCM	3.988%	5.450%	9.016%
FM-AFCM	4.171%	4.322%	5.065%
TM-AFCM	4.168%	4.322%	4.938%
EM1	6.344%	7.591%	13.768%
EM2	4.242%	5.638%	9.604%
AMRF	3.876%	4.795%	6.874%
MNI-FCM	4.979%	4.970%	5.625%
AS-FCM	3.717%	3.676%	3.832%
GradClassLeg	3.43%	4.89%	6.56%

Table 2.1: MCR results for the GradClassLeg algorithm compared with results reported in [107] for the simulated data.

2.4.3 Results on real brain data

We have applied the complete GradClassLeg algorithm to the collection of 20 complete head MRI scans from IBSR, which are provided with the expert manual segmentations. We estimate both the IIH field and the intensity class means. The Legendre polynomials are of order three. The gradient descent gain for the intensity class means estimation was set to $\alpha^\Gamma = 0.1$, and that for the IIH field was set to $\alpha^P = 0.01$. The head scan volumes are of size 256x256x64, however the registration of the data volume and the manual segmentation implies discarding some slices. Another minor inconvenient is that some other slices are severely corrupted and can not be used. Figure 2.16 shows one slice of one original MRI brain volume (Figure 2.16(a)) and the corresponding slice of the manual segmentation (Figure 2.16(b)). The slice of the restored volume after correction with the bias field estimation obtained by GradClassLeg is presented in figure 2.17(a) and the corresponding slice of the classification volume appears in figure 2.17(b). In figure 2.17(b) we show each voxel with the value estimated by GradClassLeg for the corresponding intensity class mean. As these values do not coincide in general with the visualization intensity coding values given to each class in [8], the images in figures 2.17(b) and 2.16(b) show a minor difference in brightness. We perform the classification into the GM and WM classes. To compare with results reported in other sources (i.e., [111, 148, 127, 8]), we compute the Tanimoto coefficient.

Table 2.4.3 presents the average Tanimoto coefficients for GM and WM classes over the IBSR collection, including the best result reported in the literature. The GradClassLeg algorithm proposed here performs closely to the best results re-

(a)   (b)
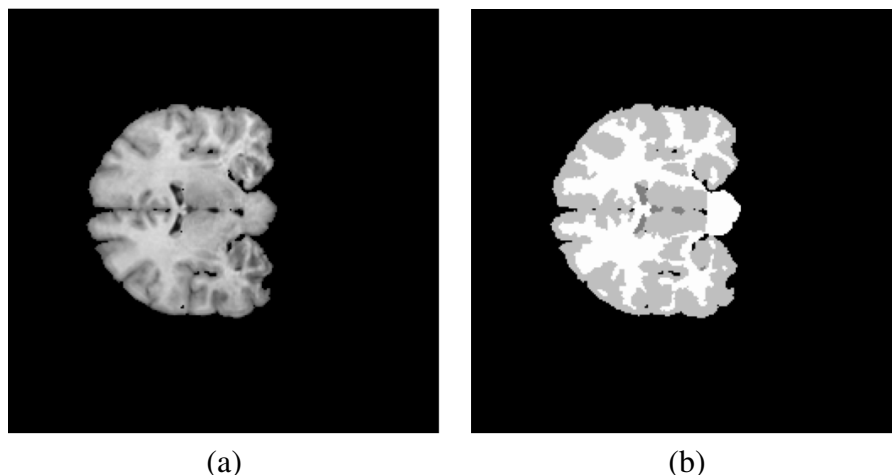
Figure 2.16: A slice from a real MRI volume: (a) original image, (b) corresponding manual segmentation. Brightest gray values correspond to WM voxels, intermediate gray values to GM voxels and darkest gray (not black) to CSF voxels.

(a)   (b)
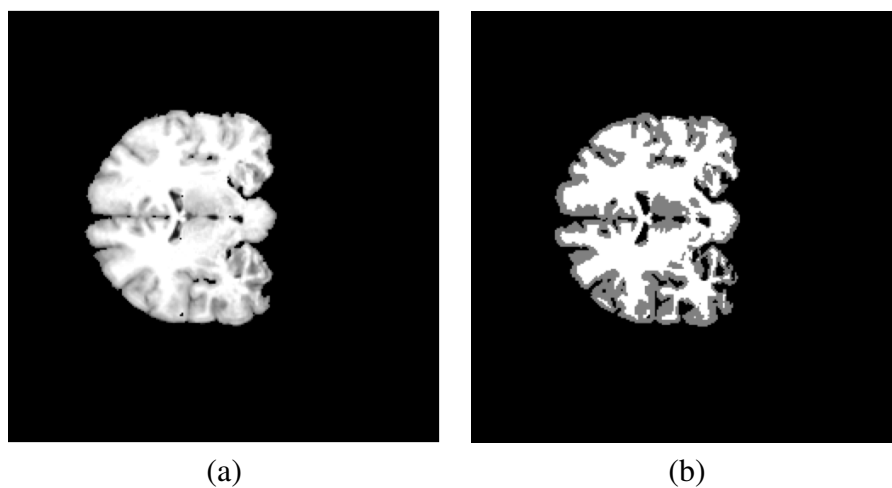
Figure 2.17: Slice of the results of GradClasLeg applied on the volume represented by the slice in figure 2.16. (a) IIH corrected image, (b) classification image. Brightest gray values correspond to WM voxels, intermediate gray values to GM voxels.

ported in the literature for this collection of MRI volumes. We have applied our algorithm to estimate the CSF along with the GM and WM intensity mean classes, performing the image classification into three classes. The accuracy decreases for the latter classes (the Tanimoto coefficients are 0.4, 0.67 and 0.66 for CSF, GM and WM classes, respectively), however these results can not be compared with the results in the literature, because all the references found reported only classification accuracies on GM and WM classes.

Algorithm	GM	WM
Wells	0.564	0.567
BMAP	0.558	0.562
BFCM	0.630	0.709
Modified FCM [148]	0.750	0.724
GradClassLeg	0.745	0.732

Table 2.2: The average Tanimoto coefficients for GM and WM classes over the IBSR collection between manual segmentation and the segmentation results of GradClassLeg, BMAP, BFCM and Modified FCM for the real data.

2.5 Computational results for AFR

The scheduling of the neighboring function width follows the expression:

$$\sigma\left(t\right) = \sigma_0 \left(\frac{\sigma_f}{\sigma_0}\right)^{t/F} \tag{2.29}$$

where σ_0 is the initial width of the neighboring function, σ_f is its final value, t is the iteration number and F is the speed of convergence to the final value of the neighboring function width: after F iterations the width is set to its final value $\forall t > F : \sigma\left(t\right) = \sigma_f$. We have used this scheduling for the SOM parameters in previous works [74, 78, 75] following some works on the optimization of k-means algorithm [39].

Sometimes we give the classification results performing the classification based on the means of the intensity classes without bias correction. We denote them BGAUSS in the tables and figures. This approach corresponds to the basic supervised classification results, which we expect to be improved upon by the bias correction and adaptive intensity class means estimation.

2.5.1 Experiments on real brain data

To start with, we apply the bare AFR to the collection of real brain datasets obtained from the IBSR.

The mean intensities were computed according to the manual segmentation. The factor F was set to $F = 1$. This value of the convergence factor implies that there will be few iterations with the final neighborhood width and that its variation will be close to linear and far from an exponential decay. The initial value of the neighboring function parameter was always set to $\sigma_0 = 30$. The number of iterations was always 20. The slices of the volumes were processed independently, and the result given is the average over all the volume.

The figures 2.18, 2.19 and 2.20 show the plot of the Tanimoto coefficients for the CSF, GM and WM for each one of the brain datasets tested. The abscissa axis labels correspond to the identification of the image in the IBSR site. The results for the AFR were obtained with a $\sigma_f = 15$. The remaining algorithm results were extracted from the IBSR site. For the CSF, the results are very good and above the results reported in the IBSR site. However, the CSF is a very scarce class that many works omit in their reports. Examination of the figures 2.19 and 2.20 show that our approach gives results comparable to the state of the art algorithms. One remarkable feature of our approach is that it gives more steady results across datasets: it maintains an average performance for GM and WM detection for some datasets for which the remaining algorithms have a big drop of performance.

The figure 2.21 shows an slice from one of the brain datasets and its manual segmentation as provided by the IBSR site. The figure 2.22 shows the estimated IIH field for this slice and the corrected image, obtained dividing the original T1-weighted image by the estimated IIH field. Figure 2.23 shows the automated segmentation of the corrected dataset. For a more precise evaluation of the sources of error, we show in figure 2.24 the voxels corresponding to each class in the original and the AFR segmented dataset. We have cropped the air around brain volumes to reduce the computational cost. It can be appreciated that a great source of error for the CSF and GM classes are the GM voxels at the contour of the brain. There the existence of voxels corresponding to air in the neighborhood introduces some boundary effects. It can also be appreciated that tissue boundaries of WM are confused with GM. Another source of error for the CSF and GM classes are the ventricles: figure 2.24(d) shows that many ventricle voxels are classified as GM, while it can be appreciated in figure 2.21(a) that their intensity is close to that of the GM voxels. This confusion suggest that *a priori* spatial information could be of great use for this process.
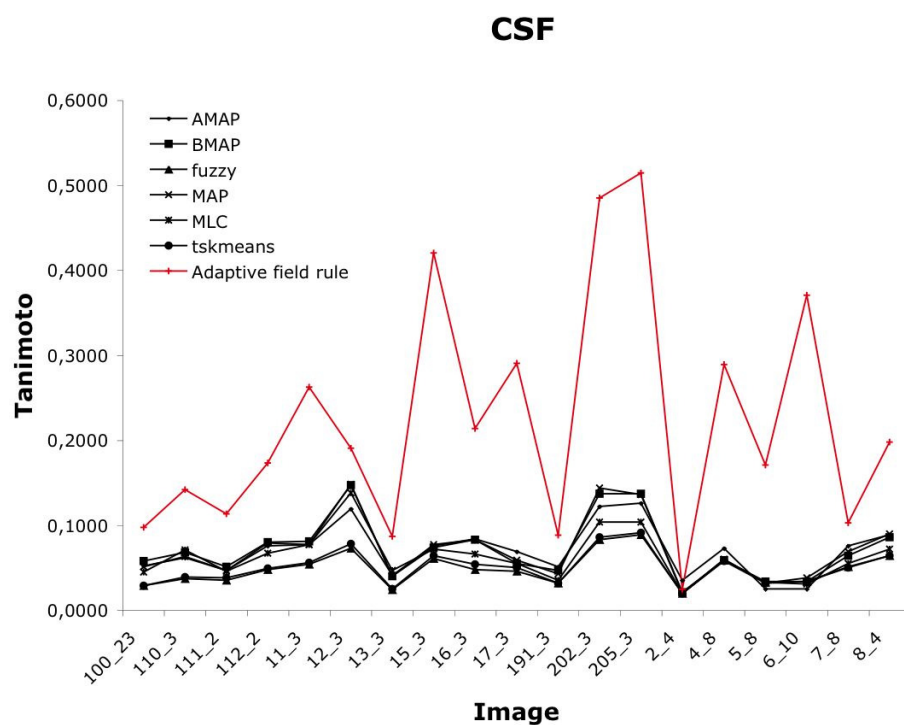
Figure 2.18: Plot of the Tanimoto coefficients of the CSF classification for all the real images treated. AFR parameters $F = 1$, $\sigma_0 = 30$, $\sigma_0 = 15$, $\alpha = 0.5$, 20 iterations.
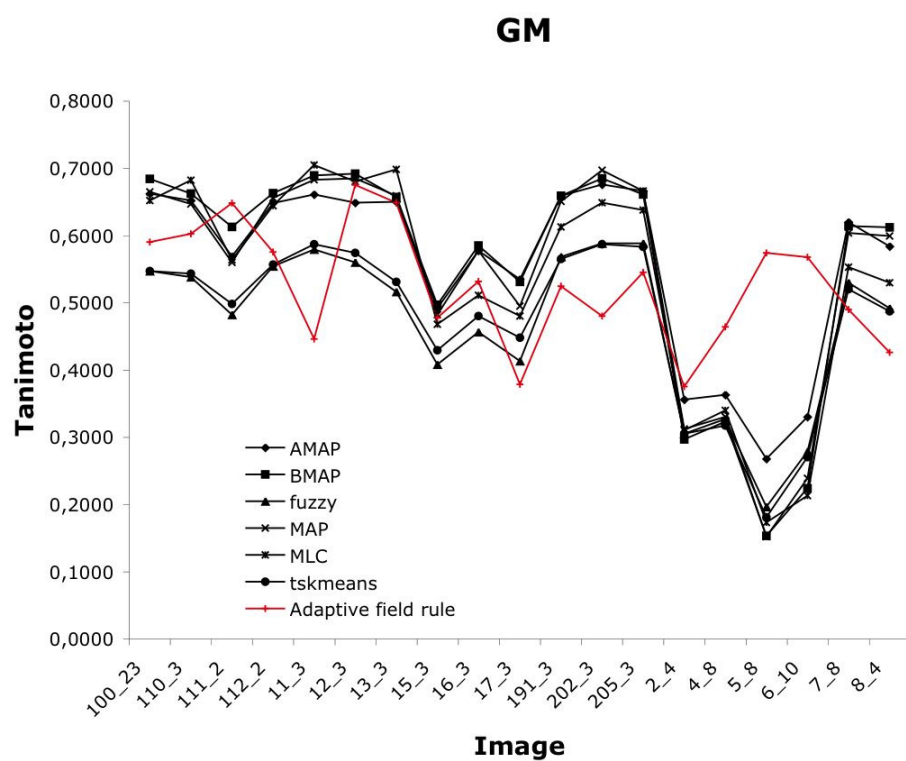
Figure 2.19: Plot of the Tanimoto coefficients of the GM classification for all the images treated. AFR parameters $F = 1$, $\sigma_0 = 30$, $\sigma_0 = 15$, $\alpha = 0.5$, 20 iterations.
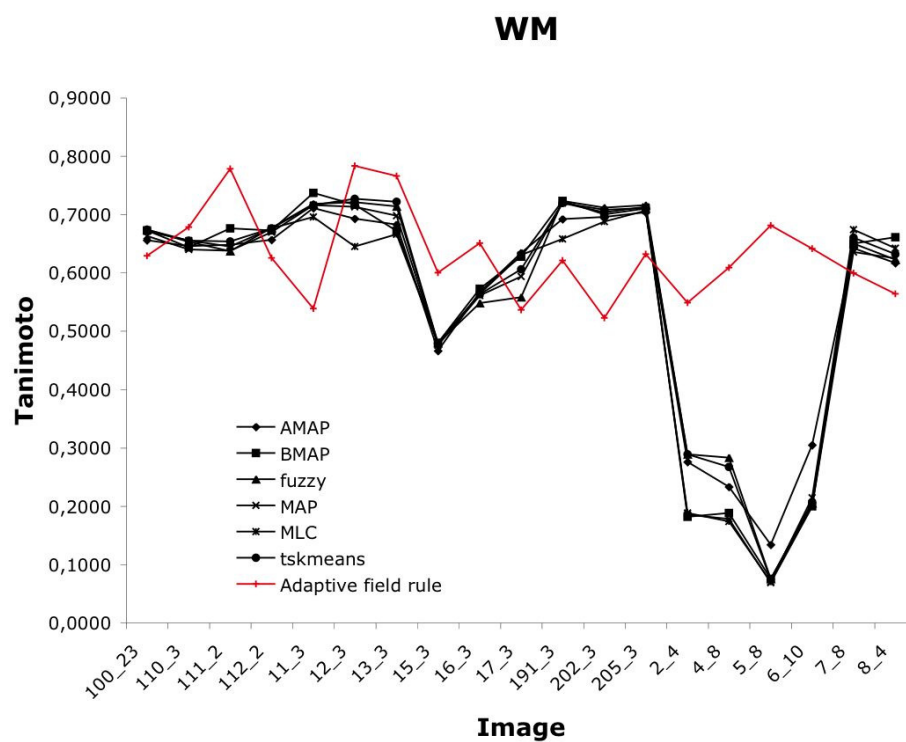
Figure 2.20: Plot of the Tanimoto coefficients of the WM classification for all the images treated. AFR parameters $F = 1$, $\sigma_0 = 30$, $\sigma_0 = 15$, $\alpha = 0.5$, 20 iterations.

Table 2.3 gives the summary Tanimoto coefficient results comparing the AFR with the state of the art results reported in the IBSR site. We have computed the means withdrawing the dataset that we were unable to match with its manual segmentation. In this table the results of the AFR were obtained with $\sigma_f = 0.01$. The results are comparable to that of the state of the art algorithms, with some improvement on the WM. Table 2.4 shows the effect of changing some of the algorithm parameters. In this table, GTM denotes that the intensity class means were computed using the manual segmentation, PM denotes that the intensity class mean values were random perturbations around the nominal ones, and ME denotes that the intensity class mean values were estimated applying one k-means step after each iteration of the bias estimation process. The reference result is the one presented in table 2.3, with $\sigma_f = 0.01$. Changing the final neighboring function standard deviation to $\sigma_f = 15$ gives some improvement on the CSF and the WM classes. This result is quite interesting, because it is against the conventional application of the neighboring shrinking scheduling. It seems that for this problem, a final neighborhood size (much) greater than the null neighborhood is more convenient to ensure smoothness and to avoid the trivial setting of the bias to the voxel residual relative to the closest class intensity mean. When we perturb the intensity class means there is a general decrease of the results. Finally, performing the estimation of the intensity class means after each iteration of the bias estimation does produce improvements on the CSF and WM classes but a decrease in the GM class.

(a)   (b)
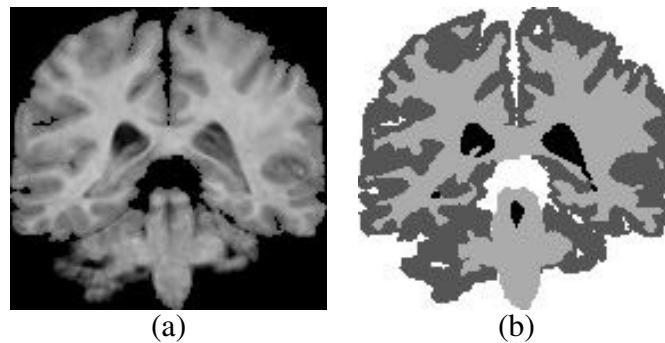
Figure 2.21: (a) Original T1-weighted coronal slice of a volume from IBSR 20 normal brain data set, and (b) its corresponding manual segmentation.

(a)   (b)
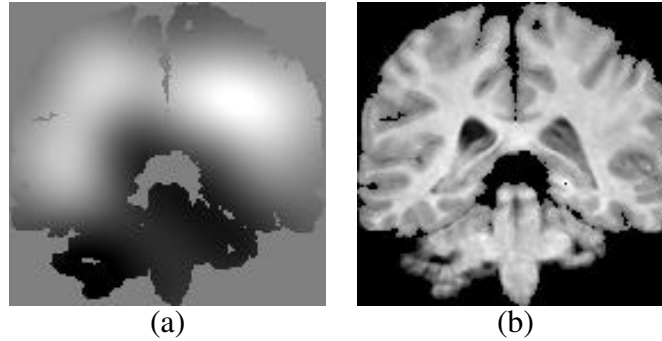
Figure 2.22: Coronal slices of the (a) estimated IIH volume and (b) IIH corrected volume, corresponding to figure 2.21(a) after AFR correction.

Figure 2.23: Segmentation corresponding to the coronal slice shown in figure 2.21(a) computed on the IIH corrected slice shown in figure 2.22(b).

Methods	CSF	GM	WM
Adaptive MAP[127]	0,0697	0,5588	0,5611
Biased MAP[127]	0.0714	0.5527	0.5559
Fuzzy c-means [18]	0.0484	0.4698	0.5608
Maximum A Posteriori Probability (MAP)	0.0714	0.5452	0.5473
Maximum-Likelihood (MLC)	0.0631	0.5317	0.5444
Tree-structure k-means	0.0499	0.4742	0.5653
Adaptive Field Rule (AFR)	0.0918	0.5570	0.5867

Table 2.3: Summary Tanimoto coefficient results for the AFR compared with results, from the IBSR site, of various algorithms.
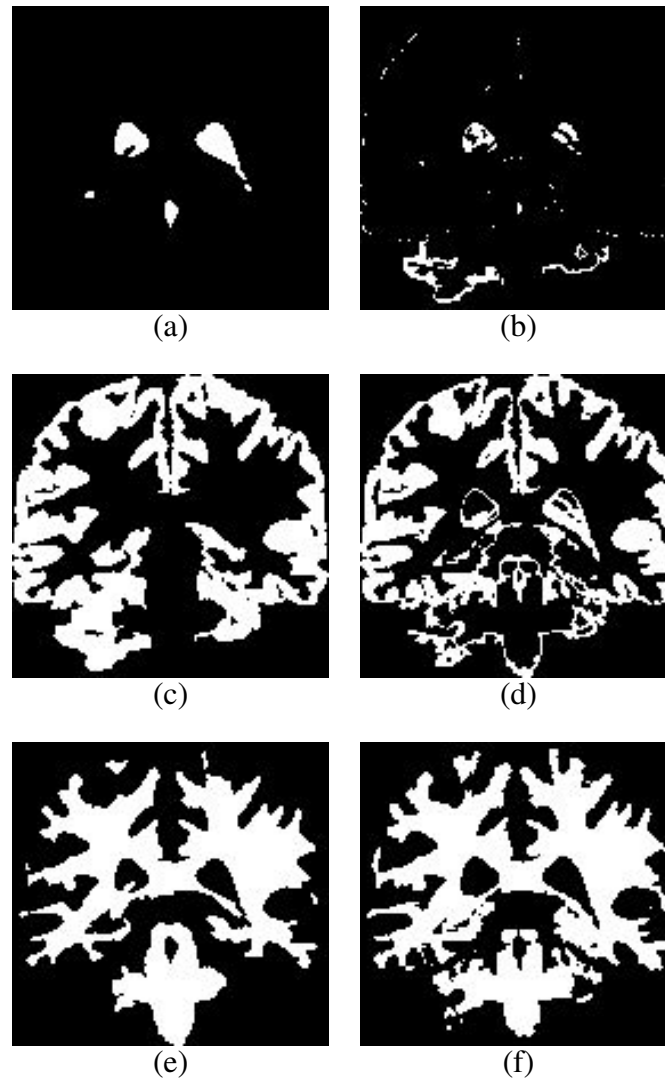
Figure 2.24: Manual segmentation of the coronal slice shown in figure 2.21(a) into (a) CSF , (c) GM and (e) WM, and corresponding segmentation after AFR IIH correction into (b) CSF, (d) GM and (f) WM, extracted from the image in figure 2.23.

AFR settings	CSF	GM	WM
$\sigma_f = 0.01$, GTM	0.0918	0.5570	0.5867
$\sigma_f = 15$, GTM	0.1653	0.5427	0.6474
$\sigma_f = 0.01$, PM	0.1838	0.5286	0.6072
$\sigma_f = 15$, PM	0.1561	0.5436	0.6503
$\sigma_f = 0.01$, ME	0.2121	0.4981	0.6153
$\sigma_f = 15$, ME	0.2229	0.5273	0.6319

Table 2.4: Sensitivity of Tanimoto coefficient results of AFR.

2.5.2 Experiments on a simulated brain phantom

We have applied the bare AFR and the unsupervised mean estimation algorithm AFR-U to the segmentation of two simulated brain MRI volumes obtained from the BrainWeb. For the bare AFR, the intensity class means are computed as the intensity class means according to the manual segmentation. When we perform the intensity class means unsupervised estimation, applying AFR-U as defined by equation (2.24) the initial values are computed with the k-means algorithm. The factor F was set to $F = 1$. The initial value of the neighboring function parameter was set to $\sigma_0 = 30$ for the images with 40% IIH and to $\sigma_0 = 15$ for the images with 20%IIH. The number of iterations was 30.

The table 2.5 presents the results of the basic supervised Gaussian classifier (BGAUSS), the AFR and the unsupervised AFR-U. Each table entry contains the triplet of Tanimoto coefficients for the CSF, GM and WM tissue classes. It can be appreciated that the proposed approach improves over the BGAUSS classifier, and that this improvement is more clear for the strongest IIH field, that is, improvements are greater for the 40% IIH volumes than for the 20% IIH volumes.

Besides, the unsupervised intensity mean estimation computed with AFR-U introduces further improvements. The Tanimoto coefficient for the CSF follows a curious pattern, it is lower in the results obtained by the AFR and AFR-U algorithms than in the BGAUSS supervised approach for low additive noise levels. However, as the additive noise increases, the results are better for AFR and AFR-U than for BGAUSS.

Figure 2.25 shows the classification results for a coronal slice of the 40% IIH brain volume. The figure 2.25(a) shows the original slice image extracted from the volume. The figure 2.25(b) shows the distribution of CSF, GM and WM in the anatomic model used for the generation of the simulated data which is the gold standard for the image segmentation. This information has been used as a mask to

perform the brain extraction to obtain the brain shown in figure 2.25(a). The result of classification after IIH correction with the bias estimated by the AFR algorithm is shown in figure 2.25(c). The result of slice classification into CSF, GM and WM after intensity class means estimation and IIH correction by the AFR-U algorithm is shown in figure 2.25(d). There are very small differences between both AFR and AFR-U results and, between them and the gold standard.

%IIH, %noise	BGAUSS	AFR	AFR-U
20, 0	$(0.95, 0.88, 0.92)$	$(0.94, 0.89, 0.92)$	$(0.94, 0.90, 0.94)$
20, 3	$(0.94, 0.85, 0.90)$	$(0.93, 0.84, 0.92)$	$(0.93, 0.89, 0.93)$
20, 5	$(0.86, 0.80, 0.86)$	$(0.87, 0.82, 0.88)$	$(0, 88, 0.86, 0.90)$
20, 7	$(0.81, 0.73, 0.81)$	$(0.82, 0.79, 0.85)$	$(0.83, 0.80, 0.86)$
20, 9	$(0.76, 0.66, 0.77)$	$(0.78, 0.70, 0.79)$	$(0.80, 0.76, 0.80)$
40, 0	$(0.91, 0.79, 0.85)$	$(0.89, 0.83, 0.90)$	$(0.89, 0.85, 0.92)$
40, 3	$(0.90, 0.78, 084)$	$(0.89, 0.81, 0.89)$	$(0.88, 0.83, 0.91)$
40, 5	$(0.84, 0.74, 0.82)$	$(0.88, 0.77, 0.86)$	$(0.86, 0.77, 0.87)$
40, 7	$(0.78, 0.65, 0.79)$	$(0.82, 0.67, 0.81)$	$(0.82, 0.67, 0.81)$
40, 9	$(0.73, 0.55, 0.75)$	$(0.75, 0.59, 0.75)$	$(0.76, 0.59, 0.75)$

Table 2.5: Tanimoto coefficients for CSF, GM and WM tissue classes obtained on the BrainWeb. The simulated phantom has been corrupted with additive noise and IIH.

2.5.3   Some results on intensity class means estimation

To explore the simultaneous estimation of the IIH field and the intensity class means we have used the IBSR V2.0 collection of brain volumes, the 1.5mm data distribution, which is more recent than the collection used for the experiments in section 2.5.1. The data description mentions that an intensity inhomogeneity correction process has been realized on the data, therefore the expected results of AFR are the local features like partial volumes.

In figure 2.26 we have coronal slices of two brain MRI volumes from the IBSR V2.0 collection. It must be noted that they have been normalized in position, but that their sizes are quite different, they have not been subjected to nonlinear registration to a common template or atlas references. Their respective ground truth for classification are given by the manual segmentation into CSF, GM and WM shown in figure 2.27. A coronal slice in the same position as the ones shown in figure

(a)

(b)

(c)

(d)
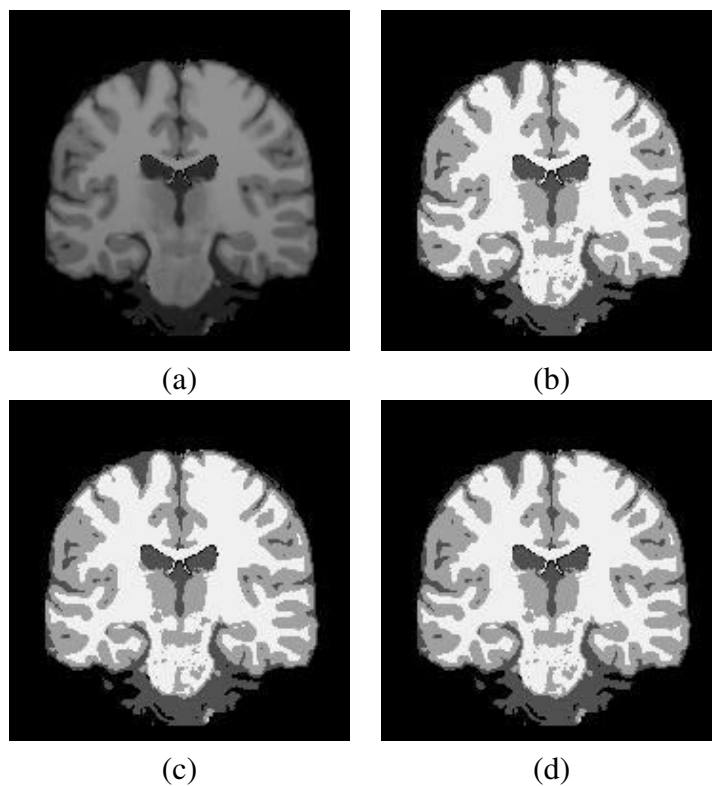
Figure 2.25: Classification results for one coronal slice of the simulated brain phantom volume with 40% Intensity Inhomogeneity from the BrainWeb site. (a) original skull stripped slice, (b) tissue distribution in the anatomic model providing the classification ground truth, (c) classification after IIH estimation and correction with AFR, (d) classification after IIH and intensity class means estimation with AFR-U.

2.26 of the crisp atlas used by the algorithm AFR-A of equation (2.25), obtained from the image manual segmentations by majority voting on each voxel, is shown in figure 2.28. The differences in brain sizes produce an extended GM region in the crisp atlas, covering the GM regions of both small and bigger brains. The probabilistic atlases for CSF, GM and WM, estimated from the volume manual segmentations, used by AFR-PA of equation (2.26) are presented in figure 2.29. The variability of brain sizes produces a thick strip in the probabilistic GM atlas, and a star like shape in the probabilistic WM atlas. Finally, figure 2.30 shows the segmentation results obtained with AFR-U, AFR-A, AFR-PA and AFR-S from the k-means initialization.

(a)   (b)
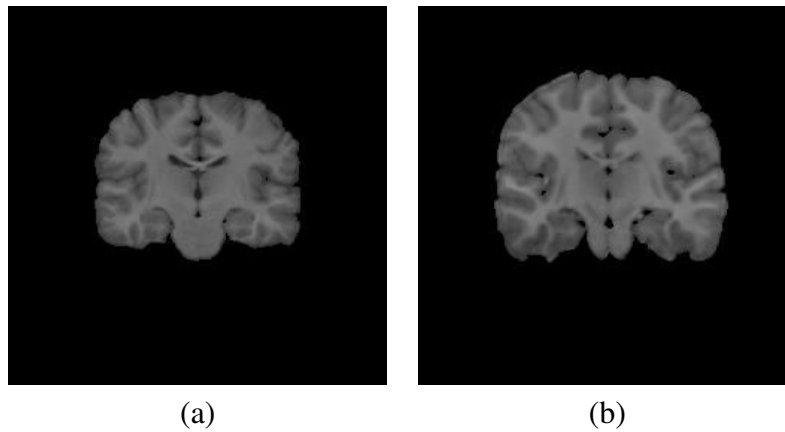
Figure 2.26: Original coronal slices of real MRI brain volumes from IBSR V2.0 collection.

Table 2.6 shows the average Tanimoto coefficients for the CSF, GM and WM over the collection of brain volumes. We call BGAUSS the results obtained with the Basic Gaussian classifier using the manual segmentation of each volume as the ground truth to compute the tissue means. The GT denote that the AFR variants are initialized with those means while KM denotes that the algorithms start from the means estimated by the k-means algorithm. The main conclusion that can be obtained from the table is that results are very poor in general. The standard application of AFR, assuming the manual segmentation intensity class means, barely improves the Gaussian supervised classifier. That can be due to the data being already intensity normalized. The AFR assuming the k-means initialization is very poor for the CSF and GM classes. The unsupervised estimation of the means performed by AFR-U does not improve the results any further. The crisp atlas used by the AFR-A seems to be of some help but the results do not improve

(a)   (b)
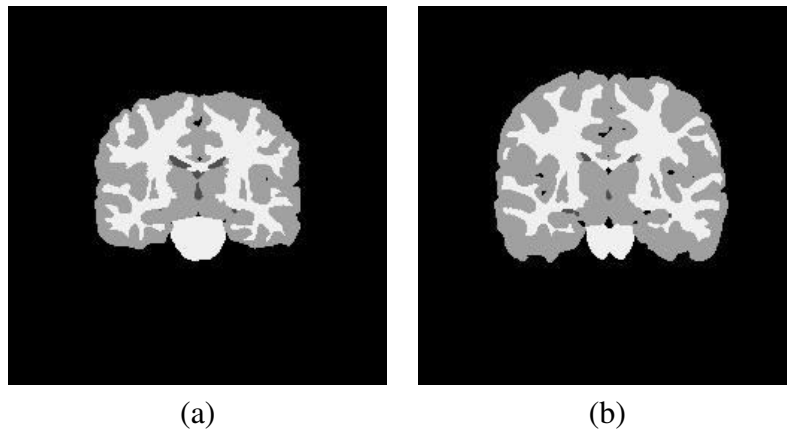
Figure 2.27: Manual segmentation into CSF, GM and WM of the slice images in figure 2.26.
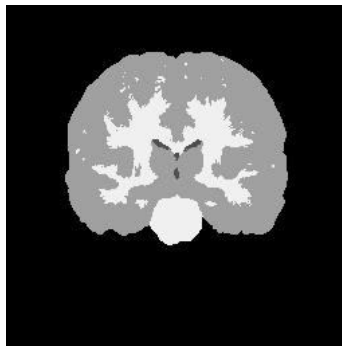
Figure 2.28: Crisp atlas obtained from the manual segmentations of the IBSR v2.0 collection, by majority voting at each voxel. Coronal slice in the same position as those in figure 2.26.
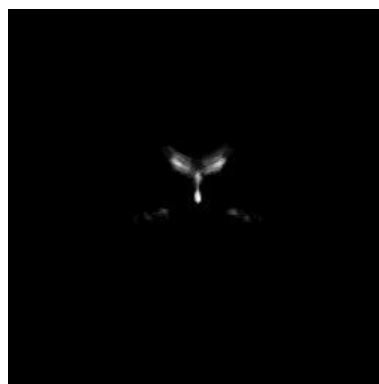
(a)
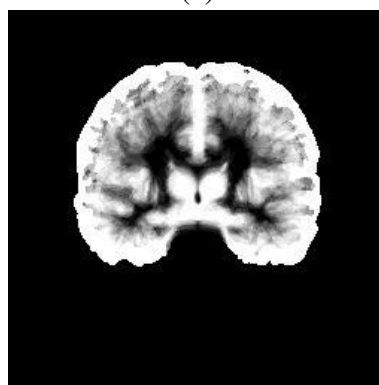
(b)
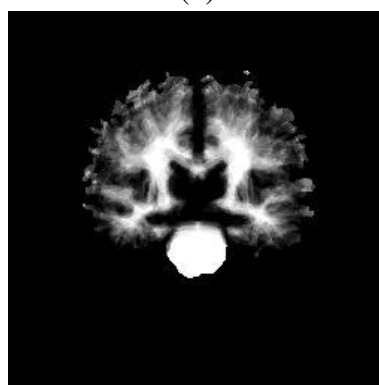
(c)

Figure 2.29: Probabilistic atlases for (a) CSF, (b) GM and (c) WM computed from the manual segmentation of the volumes in the IBSR v2.0 collection. Coronal slices corresponding to those in figure 2.26.
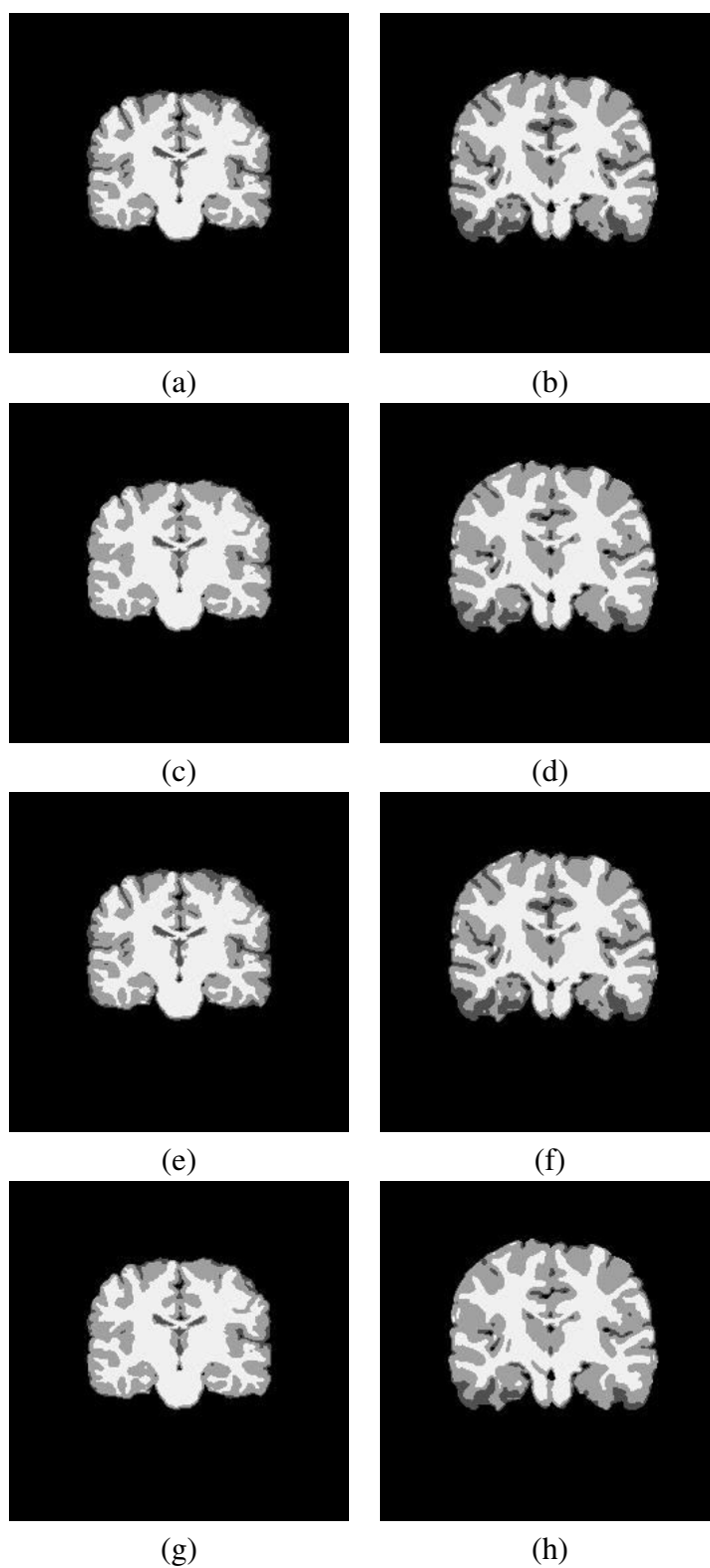
Figure 2.30: Classification into CSF, GM and WM by the (a,b) AFR-U, (c,d) AFR-A, (e,f) AFR-PA and (g,h) AFR-S. Coronal slices corresponding to the ones shown in figure 2.26.

much on the bare AFR with ground truth initialization. It can be interpreted as the AFR-A being able to use collective *a priori* spatial distribution information to obtain results comparable to the knowledge of the manual segmentation for the computation of the intensity class means. The AFR-PA that uses the spatial information from the probabilistic atlas, on the other hand, does not give good results (and was tricky to tune to convergence). Finally, the supervised approach does not give the expected improvement.

The use of spatial distribution atlases holds many promises, but also hides many traps. The unsupervised segmentation of the images has severe limits, as many voxels from one tissue will have intensity values close to that of other classes. Having some spatial information about the distribution of tissues seems the sure way to solve the problem. But to compute this spatial distribution and to obtain the desired atlas we need either to have a way to spatially normalize the volumes without introducing much distortion, or to select very precisely the population of images from which the atlas will be computed. In the case at hand, the variation in brain size introduces some undesirable structures in the crisp and probabilistic atlases, because we have built them without taking into account that the size conditions the spatial distribution. Templates reported in the literature are built on the intensity domain, not upon the manual segmentations. Warping classification images is a technique unheard of. Performing linear scale transformation may pose problems, because we are dealing with discrete class information. Approaches reported in the literature that use some kind of template or atlas, need to be taken with a precise definition of the population the atlas was derived from.

The examination of the manual segmentation gives an explanation for these results. Comparison of the segmentations shown in figure 2.27 with the BrainWeb model shown in figure 2.25(b) allows to appreciate that some inner brain regions that are classified as WM in the BrainWeb model are systematically classified as GM in the IBSR manual segmentations. The intensity values in these regions seem to correspond to WM voxels. We think that there is a systematic error in the manual segmentation that is reflected in the very low values of BGAUSS and the supervised AFR-S entries in table 2.6.

2.6 Conclusions

Because the PhD Thesis works reported in this chapter deal with two quite different and complementary approaches and results are rather disjoint, we will separate the conclussions into two subsections devoted to each approach. Although it has

	CSF	GM	WM
BGAUSS	0.09	0.62	0.70
AFR GT	0.12	0.62	0.71
AFR KM	0.06	0.52	0.71
AFR-U GT	0.10	0.58	0.68
AFR-U KM	0.06	0.52	0.71
AFR-A KM	0.11	0.62	0.71
AFR-PA KM	0.07	0.52	0.72
AFR-S GT	0.16	0.63	0.67
AFR-S KM	0.12	0.64	0.71

Table 2.6: Average Tanimoto coefficients over the IBSR V2.0 collection of brain volumes of intensity class mean estimation approaches combined with AFR.

been commented several times along the chapter, we feel necessary to emphasize here the complementary nature of both approaches. In summary, GradClassLeg parametric modeling approach makes it better suited for problems with global smooth variations of the IIH, while the non-parametric nature of AFR makes it better suited for the detection of local IIH features in MRI data, such as partial volume effects.

2.6.1   GradClassLeg

We have presented a parametric gradient descent IIH correction algorithm called GradClassLeg, which assumes that the IIH field model is a linear combination of outer products of 1D Legendre polynomials. Our approach estimates both the IIH model parameters and the intensity class means. We have discussed the advantages of the proposed algorithm and the conditions for its convergence to the global minima. The GradClassLeg has been tested and compared with three other state of the art algorithms found in the literature: BMAP, Wells' EM algorithm and the BFCM algorithm. The intensity restoration results of GradClassLeg over a set of MRI images corrupted with simulated IIH fields, synthesized as random products of Legendre polynomials, confirm that the approach is robust enough to estimate strong IIH fields. Further work can be addressed to computational experiments with other synthetic IIH field models to assess these results, including more realistic IIH models. In this chapter we have show that GradClassLeg can be very robust against bad initial estimations of the intensity class means.

These experiments will be extended using other real life images available, with the inconvenience of the uncertainty added by the manual segmentations. The robustness of the algorithm makes it a good candidate to explore situations with less *a priori* information, when there are anomalies (i.e., tumorous tissues) where unsupervised segmentation can be useful to detect the unknown image structures.

On the computation time side, GradClassLeg can be faster than the other algorithms because the gradient computation is faster than the convolutions required by the other algorithms, if the basis of products of Legendre polynomials are precomputed. The GradClassLeg also performed better than the other algorithms over a set of real brain scans obtained from the IBSR, giving state of the art results.

GradClassLeg demonstrates the power of parametric approaches to IIH correction in MRI images. However, the classification component of this algorithm is rather naive. It can be expected that the results could improve if the proposed parametric IIH field estimation is combined with more sophisticated classification schemes, allowing even for the detection of the number of tissue classes in the image, a feature desirable for the detection of anomalous or pathologic conditions. This direction seems a natural and promising path to continue the PhD Thesis works reported in this chapter.

2.6.2 AFR

We have proposed an Adaptive Field Rule (AFR), which is the gradient descent of an energy function, for IIH field estimation in MRI. Our approach is based on a topological preservation formulation of the smoothness constraint on the IIH field. The proposed estimation rule is very similar to the SOM rule. The steady states of the adaptive rule satisfy some smoothness characterizations that allow assuming them as appropriate representations of the IIH fields in the image, minimizing the effect of partial volume and other sources of noise in the estimation of the IIH field. We have tested the approach on a benchmark set of real life brain images. The results show that AFR gives state of the art results, under the assumption of the knowledge of the intensity class means. We have proposed and discussed several ways to perform the estimation of the intensity class means along with the AFR IIH field estimation. On a simulated brain phantom we obtain better results with stronger intensity inhomogeneities. Combining AFR with the unsupervised learning of the intensity class means we consistently improve the AFR results on this brain phantom. When testing the use of spatial distribution atlases on the newest IBSR collection of brain datasets, computed from the man-

ual segmentations, and the supervised learning the results are quite inconsistent, and disappointing. The positive result is that we find that the use of the crisp atlas by AFR-A can give results comparable to the Gaussian supervised classifier. On a closer inspection of both results and manual segmentation images, we find that the manual segmentation published in the ISBR site can be a source of error, both for the construction of supervised classifiers and for the evaluation of supervised and unsupervised approaches. Some of the regions that are consistently classified as Gray Matter in the IBSR manual segmentations correspond to White Matter in comparable segmentations of the BrainWeb phantom. This may, at least partially, account for the poor results found by our attempts to apply AFR and the intensity class means estimation algorithms to this data.

Chapter 3

On the detection of Alzheimer's disease

In this chapter we will report on the works we have done to evaluate the feasibility of the early detection of Alzheimer's disease (AD) on the basis of the analysis of structural MRI brain data using Machine Learning tools. In summary, we use morphometric tools to select the brain regions that may serve us to extract the features upon which the classifiers could be built.

We will introduce the motivation of the work in section 3.1, including a review of the approaches that we have found in the literature dealing with this problem. In section 3.2 we give a summarized description of the process followed. Section 3.3 gives a description of the subjects under study. Feature extraction details are explained in section 3.4. Section 3.5 contains a brief overview of the classification systems used in this study, while a more formal definition has been collected in Appendix B for the sake of completeness. We discuss classifier performance indices for validation in section 3.6. Computational results are given in section 3.7.

3.1 Motivation of the work

Alzheimer's disease (AD) is a neurodegenerative disorder, which is one of the most common cause of dementia in old people. Currently, due to the socio-economic importance of the disease in western countries it is one of the most studied. The diagnosis of AD is done after the exclusion of other forms of dementia but definitive diagnosis can only be made after a post-mortem study of the

brain tissue. This is one of the reasons why MRI based early diagnosis is a current research goal in the neurosciences.

There are studies applying Support Vector Machine (SVM) either with linear [128, 177, 93, 165] or nonlinear [53, 101] kernels, to discriminate AD patients from controls based on Positron Emission Tomography (PET), Single-Photon Emission Tomography (SPECT) functional volumes [128, 140, 64, 19] or studies that combine structural and functional information such as [52], where sMRI and PET volumes are used.

The use of Artificial Neural Networks (ANNs) and Voxel-Based Morphometry (VBM) for AD detection have been reported in [86], where a single three-layer, feed-forward ANN trained with a backpropagation algorithm was used as a classifier over a small set of unpublished proprietary sMRI data. They perform data dimensionality reduction applying a PCA to improve the efficiency of the classifier. Also in [47] ANNs (multilayer perceptrons, polinomial nets and Kohonen LVQ classifiers) are used, but in this case they have analyzed three Diffussion weighted MR (DWI) images for the evaluation of CSF volume and its correlation with the advance of AD.

There are different ways to extract features from MRI for SVM classification: based on morphometric methods [43, 53, 177, 86], based on ROIs/VOIs (regions-of-interest/volumes-of-interest) [109, 101, 64] or GM voxels in automated segmentation images [93]. There are also studies to understand the improvement in the SVM classifier by adding covariates such as demographic or genotype information [165]. Work has also been reported on the selection of a small set of the most informative features for classification, such as the SVM-Recursive Feature Elimination [53], the selection based on statistical tests [109, 128], the wavelet decomposition of the RAVENS maps [101], among others. The approach applied in this study is to use the VBM detected clusters of significant voxels as a mask on the MRI and GM segmentation images to select the potentially most discriminating voxels. Feature vectors for classification are either the voxel values or some summary statistics computed over the voxels of each cluster. We both consider the feature vector computed from all the VBM clusters and the combination of the individual classifiers built from the clusters independently.

3.2   Summarized description of the process

Morphometry analysis has become a common tool for computational brain anatomy studies. It allows a comprehensive measurement of structural differences within

or across groups, not only in specific structures but throughout the entire brain. Voxel-based morphometry (VBM) is a computational approach to neuroanatomy that measures differences in local concentrations of brain tissue, through a voxel-wise comparison of multiple brain images [20]. The procedure involves the spatial normalization of subject images into a standard space, segmentation of tissue classes using *a priori* probability maps, smoothing to correct noise and small variations, and voxel-wise statistical tests. Statistical analysis is based on the General Linear Model (GLM) to describe the data in terms of experimental and confounding effects, and residual variability. Classical statistical inference is used to test hypotheses that are expressed in terms of GLM estimated regression parameters. The computation of a given contrast provides a Statistical Parametric Map, which is thresholded according to the Random Field Theory. The result of the SPM analysis for VBM is the identification of clusters of voxels that show significant effects.

Machine Learning methods have become very popular to classify functional or structural brain images to discriminate them into two classes: normal subjects or subjects with a specific neurodegenerative disorder. To solve this task we have applied:

- Support Vector Machine (SVM) classifiers with a linear and a nonlinear (RBF) kernel

- Artificial Neural Networks (ANN) classifiers

  - The Backpropagation of errors algorithm (BP) training the Feedforward Networks (sometimes called Multilayer Perceptron),

  - Radial Basis Networks (RBF),

  - Learning Vector Quantization (LVQ) and

  - Probabilistic Neural Networks (PNN).

- Combinations of classifiers

  - Simple majority voting,

  - Weighted individual SVM

  - Adaptive Boosting (AdaBoost).

A flow chart of the whole detection process is shown in figure 3.1. The VBM process is illustrated in the box embedded in the figure. The sMRI volumes are spatially normalized, then the GM is segmented, the resulting segmentation image is modulated to account for effects of the spatial normalization and smoothed, then the General Linear Model (GLM) is solved with a design matrix that may include some covariates besides the indicative variables, and, finally, statistical inference is applied to obtain the Statistical Parametric Map and the clusters of significant voxels. More detail on VBM is given in Appendix D. The VBM is applied to the set of sMRI volumes used for the experiment. The locus of clusters of significant voxels are used to extract classification feature vectors which are used to train and validate the classification systems build with the above mentioned Machine Learning techniques.

3.3 Subjects

Many of the classification studies on the detection of AD were done over populations mixing men and women. However, this mixture of populations does not seem wise as it has been demonstrated that brains of women are different from men's to the extent that it is possible to discriminate the gender via MRI analysis [101]. Moreover, it has been shown that VBM is sensitive to the gender differences. For these reasons, we have been very cautious in the selection of these subjects for this study. We have selected a set of 98 MRI women's brain volumes. It must be noted that this is a large number of subjects compared with many of the studies referred in the state of the art published works.

Ninety eight right-handed women (aged 65-96 yr) were selected from the Open Access Series of Imaging Studies (OASIS) database [12, 110]. OASIS data set has a cross-sectional collection of 416 subjects covering the adult life span aged 18 to 96 including individuals with early-stage AD. We have ruled out a set of 200 subjects whose demographic, clinical or derived anatomic volumes information was incomplete. For the present study there are 49 subjects who have been diagnosed with very mild to mild AD and 49 nondemented. A summary of subject demographics and dementia status is shown in table 3.1.

3.3.1 Imaging Protocol

The OASIS database has been built following a strict imaging protocol, to avoid variations due to imaging protocol which would pose big image normalization
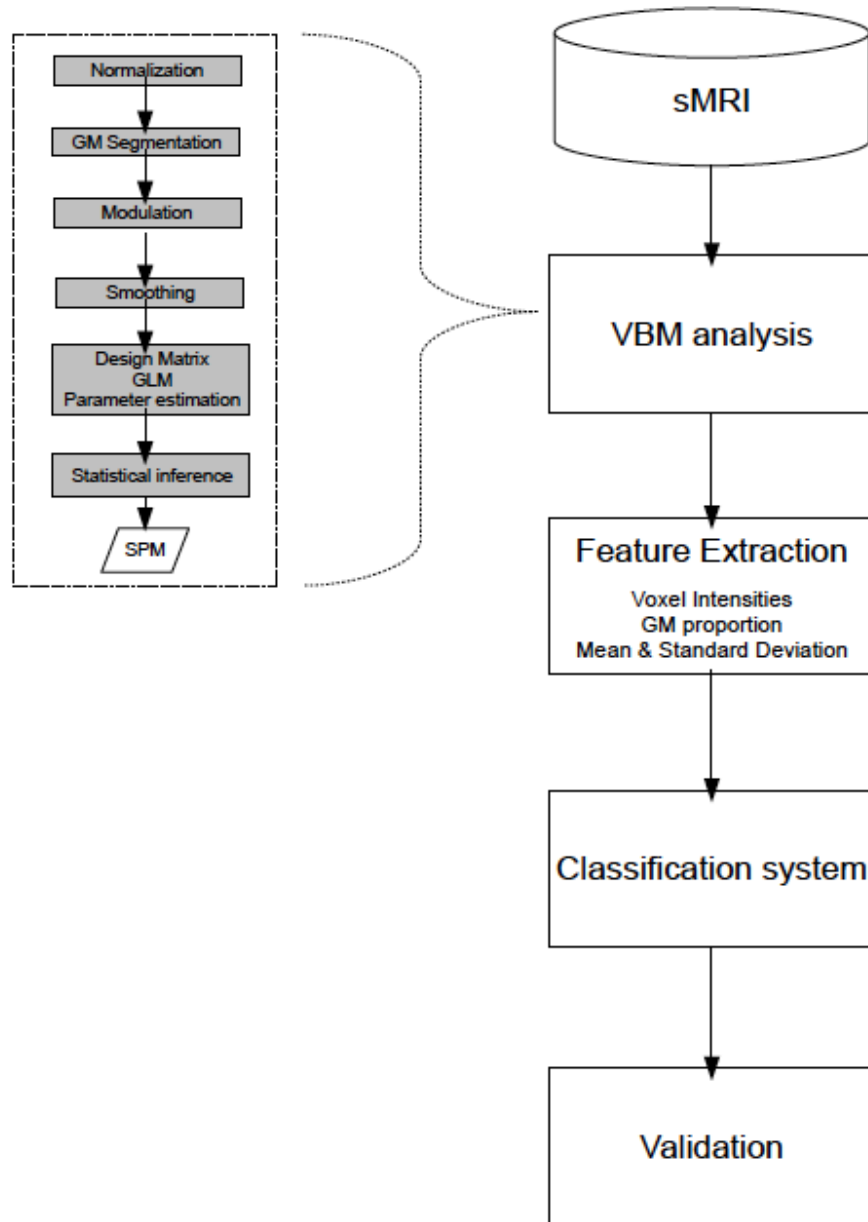
Figure 3.1: Flow chart of the classification process

	Very mild to mild AD	Normal
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5 / 1 / 2)	31 / 17 / 1	0
MMSE	24 (15-30)	28.96 (26-30)

Table 3.1: Summary of subject demographics and dementia status. Education codes correspond to the following levels of education: 1 less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best).

problems. Multiple (three or four) high-resolution structural T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) images were acquired [58] on a 1.5-T Vision scanner (Siemens, Erlangen, Germany) in a single imaging session for each subject. Image parameters: TR= 9.7 msec., TE= 4.0 msec., Flip angle= 10, TI= 20 msec., TD= 200 msec., 128 sagittal 1.25 mm slices without gaps and pixels resolution of 256×256 (1×1mm).

3.4 Feature extraction

We have tested three different feature extraction processes, based on the voxel location clusters obtained from the VBM analysis:

1. The first feature extraction process computes the ratio of GM voxels to the total number of voxels of each voxel location cluster.

2. The second feature extraction process computes the mean and standard deviation of the GM voxel intensity values of each voxel location cluster.

3. The third feature extraction process computes a very high dimensional vector with all the GM segmentation values for the voxel locations included in each VBM detected cluster. The GM segmentation voxel values were ordered in this feature vector according to the coordinate lexicographic order.

We denote these features as GMR, MSD and VV, respectively in the result tables below.

To perform the VBM analysis we have used the average MRI volume for each subject, provided in the OASIS data set. These images are already registered and re-sampled into a 1-mm isotropic image in atlas space and the bias field has been already corrected [110]. The Statistical Parametric Mapping (SPM5) [13] was used to compute the VBM which gives us the spatial mask to obtain the classification features. Images were reoriented into a right-handed coordinate system to work with SPM5. The tissue segmentation step does not need to perform bias correction. We performed the modulation normalization for GM, because we are interested in this tissue for this study. We performed a spatial smoothing before performing the voxel-wise statistics, setting the Full-Width at Half-Maximum (FWHM) of the Gaussian kernel to 10mm isotropic. A GM mask was created from the average of the GM segmentation volumes of the subjects under study. Thresholding the average GM segmentation, we obtain a binary mask that includes all voxels with probability greater than 0.1 in the average GM segmentation volume. This interpretation is not completely true, since the data is modulated, but it is close enough for the mask to be reasonable. We designed the statistical analysis as a Two-sample t-test in which the first group corresponds with AD subjects. In SPM software terms: the contrast has been set to [-1 1], a right-tailed (groupN > groupAD), correction FWE, p-value=0.05. The VBM detected clusters are used for the feature extraction for the classification procedures.

3.5 Overview of the Machine Learning Systems

In this section we comment on the Machine Learning techniques used to build the classification systems that we have evaluated for their suitability for AD detection based on the feature vectors extracted from the sMRI data. Further detail is given in Appendix B. Figure 3.2 shows a graphic overview of these classification systems. The single classifiers are the statistical SVM and the various ANN architectures. Some of these single classifiers haven used as building blocks for meta-systems that follow different strategies to combine weak single classifiers into an improved combined meta-classifier. Namely we used the nonlinear kernel SVM for this combination strategies. The two basic strategies for the combination of classifiers, one based on local classifiers for each VBM detected cluster and the other based on the construction of various global weak classifiers, have applied to the single classifiers as specified in the figure.
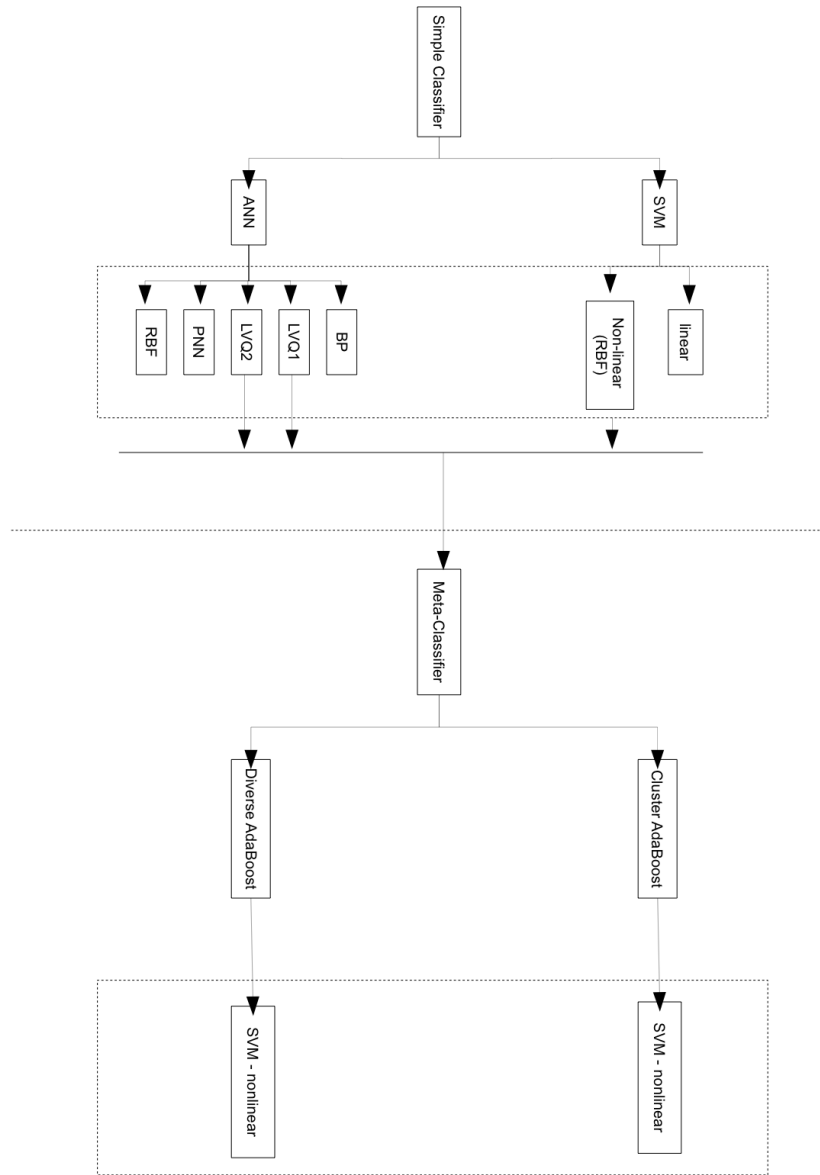
Figure 3.2: Classification Systems

3.5.1 Support Vector Machines

The Support Vector Machines (SVMs) have attracted attention from the pattern recognition community [64, 159, 129] owing to a number of theoretical and computational merits derived from [163]. SVM separates a given set of binary labelled training data with a hyperplane that is maximally distant from the two classes (known as the maximal margin hyperplane). The objective is to build a discriminating function using training data that will correctly classify new examples (x, y). When no linear separation of the training data is possible, SVMs can work effectively in combination with kernel techniques using the kernel trick, so that the hyperplane defining the SVMs corresponds to a nonlinear decision boundary in the input space that is mapped to a linearised higher- dimensional space [163]. In this way the decision function can be expressed in terms of the support vectors only:

$$f\left(\mathbf{x}\right) = sign\left(\sum \alpha_i y_i K\left(\mathbf{s}_i, \mathbf{x}\right) + w_0\right)$$

where $K(.,.)$ is a kernel function, α_i is a weight constant derived from the SVM process and the $\mathbf{s}_i$ are the support vectors [163].

3.5.2 Artificial Neural Networks

The Artificial Neural Networks (ANNs) applied in this study were the Feedforward Networks (sometimes called Multilayer Perceptron) trained with the Backpropagation of errors algorithm (BP), Radial Basis Networks (RBF), Learning Vector Quantization (LVQ) and Probabilistic Neural Networks (PNN).

Backpropagation

Backward propagation of errors or Backpropagation (BP) [138, 83, 82] is a supervised learning method. It is a nonlinear generalization of the squared error gradient descent learning rule for updating the weights of the artificial neurons in a single-layer perceptron, generalized to feed-forward networks. BP requires that the activation function used by the artificial neurons (or "nodes") is differentiable with its derivative being a simple function of itself. The BP allows to compute the gradient of the error function relative to the hidden units, and, thus, to perform the hidden unit weight estimation by gradient descent in the same manner as it is performed for the output units. Backpropagated errors are analytically derived using the classical calculus chain rule. In on-line learning the weights of the network

are updated at each input data item presentation. We have used the resilient back-propagation version, which uses only the derivative sign to perform the weight updating.

Radial Basis Function Networks

Radial Basis Function (RBF) networks [38] are ANNs that use RBFs as activation functions. RBF consist of a two layer neural network, where each hidden unit implements a radial activated function. The output units compute a weighted sum of hidden unit outputs. Training consists of the unsupervised training of the hidden units followed by the supervised training of the output units' weights. Unsupervised training of the hidden units can be performed by any clustering method, including k-means, so that many alternatives have beed proposed in the literature. The training of the output units weights is a simple linear gradient descent on the error function.

Probabilistic Neural Networks

A Probabilistic Neural Network (PNN) [153] is a special type of neural network that uses a kernel-based approximation to form an estimate of the probability density function of categories in a classification problem. To compute the response of the network to a given input, we compute distance from the input point to each of the known input points in the training sample data, and the *kernel function* is applied to the distance to compute the influence of each sample point in the response to the unknown input. The most usual kernel functions are RBFs.

Different types of RBFs could be used, but the most common is the Gaussian function. The sigma parameter of the function determines the *spread* of the RBF; that is, how quickly the function declines as the distance increases from the point. With larger sigma values the function has more spread, so that distant points have a greater influence. PNN are a kind of Nearest Neighbor classifier that uses all the data samples as reference values, the only functional transformation is the computation of the posterior probability of the classes as a combination (sum/average) of the evidence given by each data sample through its RBF window.

The tuning of a PNN network depends on selecting the optimal sigma value of the spread of the RBFs. In this paper an exhaustive search for the optimal spread value in the range $(0, 1)$ for each training set has been done. The results shown in Table 3.12 correspond to the best spread value found.

Learning Vector Quantization

Learning vector quantization (LVQ) [94, 151] provides a method for training competitive layers in a supervised manner. The system is composed of an unsupervisedly trained competitive layer which performs a partitioning of the input space. The supervisedly trained output layer provides the labeling of the input data according to its belonging to an input region (crisp clustering) or to its degree of membership (soft clustering). In the original proposition of the LVQ, the competitive units were cluster centers with the Euclidean distance as the similitude measure. Training of the competitive units can be performed by Kohonen's Self Organizing Map. Supervised training was simply the assignment of a label to a competitive unit according to a majority voting on the data samples falling in the partition corresponding to the unit.

LVQ provides fine tuning of the competitive units using class information. The basic versions proposed by Kohonen are known as the LVQ1 and LVQ2. Both start with the unsupervised learning of the competitive units, and its initial majority voting labeling.

In the LVQ1 a supervised training is performed as follows: for each data sample we compare its label with the one of its corresponding competitive unit, if the labels match (the data item is correctly classified) then the competitive unit is moved towards the input data sample, otherwise it is moved in the opposite direction. This rule may cause an unstable and oscillatory behavior if the discriminant boundary among classes is very complex.

The LVQ2 rule is proposed to improve the learning, sometimes it is recommended to apply it *after* the LVQ1. In LVQ2, for each input data sample we find the two closest competitive units. If one correctly classifies the input and the other belongs to a wrong class, and the input data lies in a window around the midplane between them, then the correct class unit is moved towards the input and the incorrect unit is moved away from the input. We have used the simplest implementations. Other variations using different similitude measures and algorithms to assign the labels to the output units, can be found in the literature.

3.5.3 Combination of single classifiers

Diverse Adaptive Boosting

Adaptive Boosting (AdaBoost) [146, 62] is a meta-algorithm for machine learning that can be used in conjunction with many other learning algorithms to improve

their performance. AdaBoost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. Otherwise, it is less susceptible to the over-fitting problem than most learning algorithms.

AdaBoost calls a weak classifier repeatedly in a series of rounds $t = 1, ..., T$. For each call a distribution of weights $W_t = \{w_i^t\}$ is updated and indicates the importance of examples in the data set for the classification. On each round, the weights of each incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on the wrongly classified examples.

Following these ideas, we have tested a combination of SVM classifiers along the ideas from the Diverse AdaBoost SVM [105], presented as Algorithm 3.1. In this approach we built a sequence of SVM classifiers of increasing variance parameter. The results of the classifiers are weighted according to their statistical error to obtain the response to the test inputs in the 10-fold validation process.

We have applied this approach only to the SVM as the single classifier building blocks.

Combination of independent classifiers built for each cluster

We have considered two ways to combine the independent classifiers constructed for each VBM detected cluster of significant voxels:

1. by a simple majority voting, and to use the cluster with greatest statistical significance to resolve ties. This can be viewed as a simplified combination of classifiers.

2. Furthermore, we have defined a combination of classifiers weighted by the individual training errors, where the classifier weights are computed as in the AdaBoostSVM algorithm [105] (we describe this combination formally in Algorithm 3.2), assuming an uniform weighting of the data samples.

We have applied this approach using the SVM as the single classifier building blocks.

3.6 Classifier performance indices

We evaluated the performance of the classifier using the 10-fold cross-validation test. To quantify the results we measured the accuracy, the ratio of the number of

---

Algorithm 3.1 Diverse AdaBoostSVM

---

1. **Input:** a set of training samples with labels $\{(x_1, y_1), \ldots, (x_N, y_N)\}$; the initial σ, σ_{ini}; the minimal σ, σ_{min}; the step of σ, σ_{step}; the threshold on diversity DIV.

2. **Initialize:** the weights of training samples: $w_i^t = 1/N$, for all $i = 1, ..., N$

3. **Do while** ($\sigma > \sigma_{ini}$)

   (a) Calculate gamma: $\gamma = (2\sigma^2)^{-1}$.

   (b) Use σ to train a component classifier h_t on the weighted training set.

   (c) Calculate the training error of h_t: $\epsilon_t = \sum_{i=1}^{N} w_i^t$, $\quad y_i \neq h_t(x_i)$.

   (d) Calculate the diversity of h_t: $D_t = \sum_{i=1}^{N} d_t(x_i)$, where $d_t(x_i) = \begin{cases} 0 & if\ h_t(x_i) = y_i \\ 1 & if\ h_t(x_i) \neq y_i \end{cases}$

   (e) Calculate the diversity of weighted component classifiers and the current classifier: $D = \sum_{t=1}^{T} \sum_{i=1}^{N} d_t(x_i)$.

   (f) If $\epsilon_t > 0.5$ or $D < DIV$: decrease σ by σ_{step} and go to (a).

   (g) Set weight of the component classifier h_t: $\alpha_t = \frac{1}{2} \ln(\frac{\epsilon_t}{1-\epsilon_t})$.

   (h) Update the weights of training samples: $w_i^{t+1} = w_i^t exp(-\alpha y_i h_t(x_i))$.

   (i) Normalize the weights of training samples: $w_i^{t+1} = w_i^{t+1} (\sum_{i=1}^{N} w_i^{t+1})^{-1}$.

4. **Output:** $f(x) = sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.

Algorithm 3.2 Combining the independent classifiers trained per cluster

---

1. **Input:** as many sets of training samples with labels as clusters in the statistical parametric map $T_k = \{(x_1, y_1), \ldots, (x_N, y_N)\}, k = 1..C$, where N is the number of samples of each cluster.

2. **Initialize:** the weights of training samples: $w_i^k = 1/N$, for all $i = 1, ..., N$

3. **For each k cluster do**

    (a) Search the best γ for the RBF kernel for the training set T_k, we denote it as γ_k.

    (b) Train the classifier with T_k and γ_k, we denote the classifier as h_k.

    (c) Classify the same training T_k set with h_k.

    (d) Calculate the training error of h_k: $\epsilon_k = \sum_{i=1}^{N} w_i^k, \quad y_i \neq h_k(x_i)$.

    (e) Compute the weight of the cluster classifier h_k: $\alpha_k = \frac{1}{2} \ln(\frac{\epsilon_k}{1-\epsilon_k})$.

4. **Output:** for each test data x its classification is $f(x) = sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.

---

test volumes correctly classified to the total of tested volumes. We also quantified the specificity and sensitivity of each test defined as

$$Specificity = \frac{TP}{TP + FP}$$

and

$$Sensitivity = \frac{TN}{TN + FN},$$

where TP is the number of true positives: number of AD patient volumes correctly classified; TN is the number of true negatives: number of control volumes correctly classified; FP is the number of false positives: number of AD patient volumes classified as control; FN is the number of false negatives: number of control volumes classified as patient.

3.7 Computational results

In this section we present the results obtained from VBM analysis and from the final validation of the classification process. A graphic overview of the classification systems is in figure 3.2, so we can follow this schema to give the final results.

3.7.1 VBM analysis

The VBM procedure involves some pre-processing steps before the statistical analysis, as it is shown in figure 3.1. Figure 3.3 shows the orthogonal views from a control and a patient before and after segmentation. The control subject is a 89 year old female with a very high education level (beyond college) and with a high-standing socioeconomic status. The patient subject is a 80 year old female, who did not finish the college and with a medium socioeconomic status.

We have performed the VBM twice, first without any covariate included in the design matrix of the GLM and second taking into account the normalized brain volume (nWBV). In the design matrix, regressors (or covariates) are additional columns. This is graphically viewed in figure 3.4, which corresponds to the design matrix when we add the nWBV covariate. Figure 3.5 shows the plot of the covariate values over transposed design matrix.

The clusters of significant voxels detected by the VBM analyses are displayed in figure 3.6. Each VBM process produces different sets of voxel location clusters,
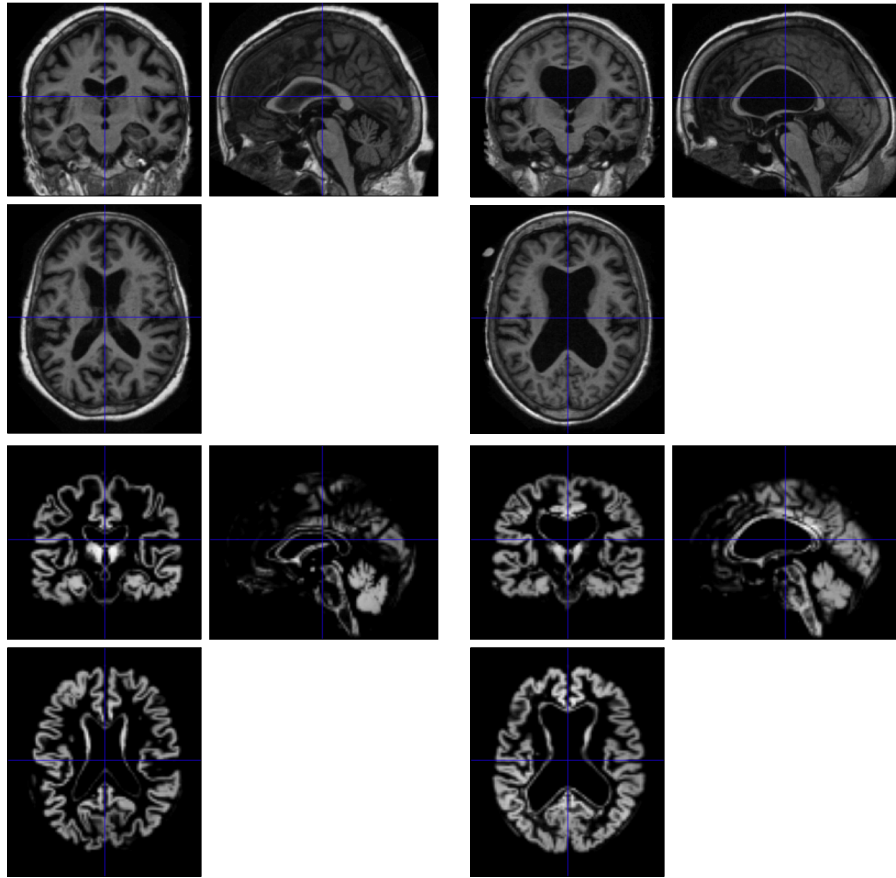
Figure 3.3: Orthogonal views of original (at the top) and segmented (at the bottom) images from a control (left column) and a patient (right column).

and, therefore, different sets of feature vectors. This information is presented in table 3.2 for the first case and 3.3 for the second one. The covariate helps to focus the VBM, giving less and smaller clusters than the VBM without covariates. Of course this implies that the feature vectors will be of lower dimensionality.
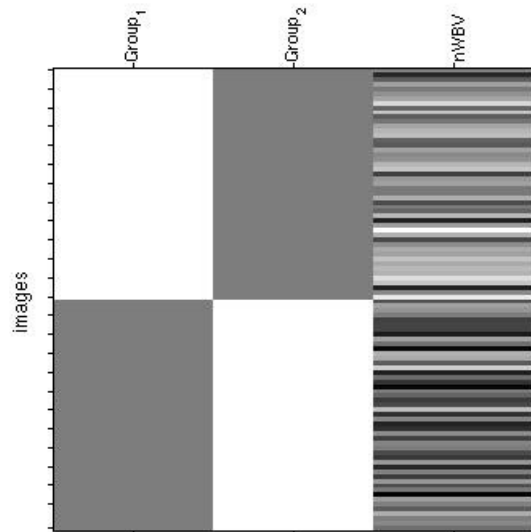
Figure 3.4: The design matrix with the nWBV covariate

3.7.2 Classification results

We present for each classification experiment the following data: the number of features, accuracy, specificity, which is related to AD patients and sensitivity, which is related to control subjects. The SVMs have been applied as single classifiers on the feature vectors obtained applying VBM with and without covariate in the design matrix. For the rest of the classifiers we have used as feature vectors those extracted from the VBM without any covariate.

Single SVM classifiers

Table 3.4 presents the results for the SVM classifiers applied to the feature vectors obtained from VBM without covariate in the design matrix. Table 3.5 presents the results for the SVM classifiers applied to the feature vectors obtained from VBM with the normalized whole-brain volume (nWBV) as the covariate in the
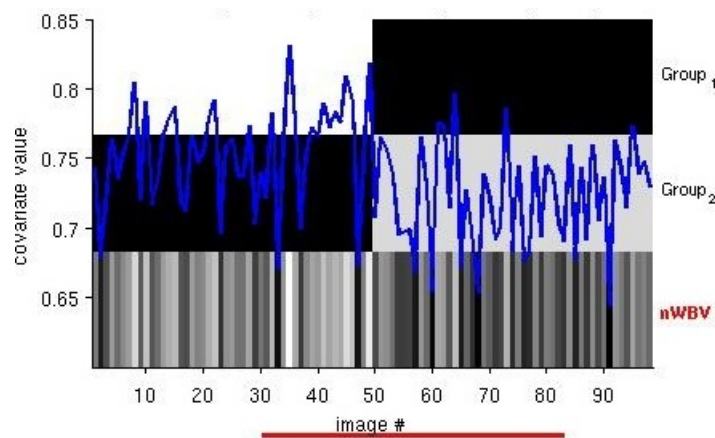
Figure 3.5: Covariate nWBV ploted over transposed design matrix
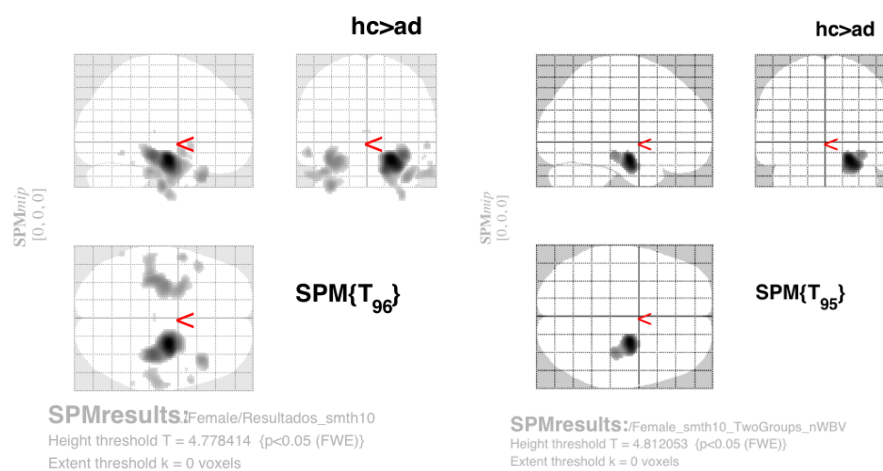
Figure 3.6: SPM results: clusters of significant voxels with increased gray matter density in the controls relative to the patient subjects, detected by the VBM process. Left: design matrix without covariate, right: nWBV included as covariate in the design matrix.

Statistics: *p−values adjusted for search volume*

set−level		cluster−level			voxel−level					mm mm mm		
p	c	$p_{corrected}$	k_E	$p_{uncorrected}$	$p_{FWE-corr}$	$p_{FDR-corr}$	T	$(Z_=)$	$p_{uncorrected}$			
0.000	12	0.000	1764	0.000	0.000	0.000	7.59	6.70	0.000	24	−8	−16
					0.000	0.000	6.43	5.85	0.000	34	−24	−12
					0.007	0.000	5.33	4.98	0.000	34	−8	−48
		0.000	1355	0.000	0.001	0.000	5.91	5.44	0.000	−34	−20	−16
					0.001	0.000	5.87	5.41	0.000	−30	−12	−18
					0.002	0.000	5.71	5.29	0.000	−34	−14	−40
		0.000	161	0.004	0.002	0.000	5.68	5.26	0.000	40	24	−30
		0.000	195	0.002	0.006	0.000	5.36	5.00	0.000	58	−20	−28
					0.018	0.000	5.07	4.76	0.000	62	−10	−22
		0.005	42	0.106	0.007	0.000	5.33	4.98	0.000	−56	4	−8
		0.003	60	0.058	0.011	0.000	5.21	4.88	0.000	−48	20	−32
		0.018	13	0.358	0.015	0.000	5.12	4.80	0.000	58	12	−2
		0.024	8	0.475	0.030	0.000	4.93	4.64	0.000	−58	−54	−10
		0.034	3	0.679	0.038	0.000	4.86	4.58	0.000	56	10	−12
		0.027	6	0.541	0.041	0.000	4.84	4.57	0.000	62	−44	−18
		0.034	3	0.679	0.043	0.000	4.82	4.55	0.000	0	−22	10
		0.042	1	0.830	0.049	0.000	4.79	4.52	0.000	48	8	−40

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 4.78, p = 0.000 (0.050) {p<0.05 (FWE)} Degrees of freedom = [1.0, 96.0]
Extent threshold: k = 0 voxels, p = 1.000 (0.050)   FWHM = 15.4 17.1 15.8 mm mm mm; 7.7 8.6 7.9 {voxels};
Expected voxels per cluster, <k> = 16.568   Volume: 1884632; 235579 voxels; 413.7 resels
Expected number of clusters, <c> = 0.05   Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 522.52 voxels)
Expected false discovery rate, <= 0.00

Table 3.2: Volume table for 'control > patient', when 0 covariates are used in the model.

Statistics: *p−values adjusted for search volume*

cluster−level			voxel−level					mm mm mm		
$p_{corrected}$	k_E	$p_{uncorrected}$	$p_{FWE-corr}$	$p_{FDR-corr}$	T	$(Z_=)$	$p_{uncorrected}$			
0.000	553	0.000	0.000	0.000	6.42	5.84	0.000	24	−8	−16
			0.010	0.000	5.27	4.92	0.000	34	−22	−12

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 4.81, p = 0.000 (0.050) {p<0.05 (FWE)} Degrees of freedom = [1.0, 95.0]
Extent threshold: k = 0 voxels, p = 1.000 (0.050)   FWHM = 15.0 16.4 15.2 mm mm mm; 7.5 8.2 7.6 {voxels};
Expected voxels per cluster, <k> = 14.525   Volume: 1884632; 235579 voxels; 462.1 resels
Expected number of clusters, <c> = 0.05   Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 467.86 voxels)
Expected false discovery rate, <= 0.00

Table 3.3: Volume table for 'control > patient', when nWBV covariate is used in the model.

Feature extracted	Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
GMR	12	69.39% / 68.36%	0.63 / 0.61	0.88 / 0.90
MSD	24	78.57% / 80.61%	0.72 / 0.75	0.88 / 0.89
VV	3611	73.47% / 76.53%	0.72 / 0.77	0.75 / 0.76

Table 3.4: Classification results with a linear kernel (lk) and a nonlinear kernel (nlk). No covariates have been taken into account in the GLM used for the VBM. The values of $\gamma = (2\sigma^2)^{-1}$ for non linear kernel were 0.5, 0.031, 0.0078 for each feature extraction process, respectively.

Feature extracted	Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
GMR	2	51% / 51%	0.50 / 0.50	1 / 1
MSD	4	69.38% / 72.45%	0.65 / 0.68	0.79 / 0.79
VV	265	66.32% / 75.51%	0.65 / 0.72	0.67 / 0.80

Table 3.5: Classification results with a linear kernel (lk) and a nonlinear kernel (nlk). The normalized whole-brain volume (nWBV) covariate has been taken into account in the GLM for the VBM. The values of γ for nlk were 0.5, 2.7, 0.004 for GMR, MSD and VV respectively.

design matrix. Each table entry contains the SVM results using the linear and RBF kernels upon the corresponding feature vector set. In both tables rows correspond to feature extraction processes as described in section 3.4.

The best accuracy result (Table 3.4) is 80.6% with the RBF kernel, but this result is not far away from the results of the linear SVM. This best accuracy result is obtained with a rather straightforward feature extraction method: the mean and standard deviation of the MRI voxel intensities. This means that MRI intensities may have discriminant value. The classification results of table 3.5, using the covariate nWBV in the design matrix of the GLM, confirm that the nonlinear SVM is more accurate. However, results in table 3.5 are systematically lower than in table 3.4.

Overall the sensitivity results in tables 3.4 and 3.5 are much lower than the specificity. We hypothesize that the source of error is the confusion of mild demented AD patients with control subjects. Mild demented AD patients are subjects with CDR=0.5 (Clinical Dementia Ratio) and a high value for the MMSE (Minimental-State Examination), i.e., MMSE=30. Therefore we repeat the feature extraction and classification experiment taking out of the population 9 mild demented AD patients. The results for the RBF kernel SVM are given in table 3.6. The classification accuracy increases from 80.6% (in the best result of table 3.4) up to 87.5%. Also sensitivity and specificity improve if we compare table 3.4

Feature extracted	Features	γ	Accuracy	Sensitivity	Specificity
GMR	12	0.9	72.5%	0.84	0.66
MSD	24	0.6	87.5%	0.89	0.86
VV	3611	1.5	86.25%	0.85	0.87

Table 3.6: Classification results of 40 AD patients vs. 49 control subjects with the SVM and a RBF kernel, 9 very mild AD subjects were taken out from the AD patients subset.

Feature extracted	Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
MSD	24	74% / 75%	0.51 / 0.56	0.97 / 0.95
VV	3611	77% / 78%	0.74 / 0.76	0.80 / 0.82

Table 3.7: Majority voting classification results with linear kernel (lk) and nonlinear kernel (nlk) SVM built independently for each VBM cluster.

and table 3.5 against table 3.6.

Results on the combination of SVM

We will give results on the simple voting of independent classifiers based on statistical significance of VBM, the weighted combination of individual cluster SVM based on training errors and an adaptive boosting strategy for combining classifiers. For all these experiments a 10-fold crossvalidation process was used.

Table 3.7 presents the results of the combination of SVM classifiers built up over each cluster independently, searching for the best kernel parameter σ in each classifier independently. The results do not improve over the ones obtained with the whole image feature vector in table 3.4. We note that, contrary to the global feature vector, the results improve when considering the whole collection of MRI voxel intensities.

Table 3.8 presents the results of the combination of individual weighted SVM classifiers. Each SVM classifier was trained with one VBM cluster feature set and the weights were computed according to its training error. We obtain a significant improvement of the accuracy when considering the voxel intensities as features for the nonlinear RBF SVM.

Table 3.9 shows the results of the Diverse AdaBoost. The σ_{min} is set as 0.1, the σ_{ini} is set as 100 and σ_{step} is set as 0.1. The DIV value is set as as 0.6. The best accuracy result is obtained for the MSD feature vector.

Feature extracted	Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
MSD	24	71% / 79%	0.54 / 0.78	0.88 / 0.80
VV	3611	73% / 86%	0.76 / 0.80	0.70 / 0.92

Table 3.8: Weighted individual SVM per cluster classification results. The value of the RBF kernels for the nonlinear (nlk) classifiers were searched for the best fit to the training set.

Feature extracted	Features	Accuracy	Sensitivity	Specificity
MSD	24	85%	0.78	0.92
VV	3611	78%	0.71	0.85

Table 3.9: Diverse AdaBoostSVM classification results.

ANN classifiers

The results shown are the values of the classification results from a 10-fold cross-validation process. We give results of the following classifiers: Backpropagation (Table 3.10), RBF (Table 3.11), PNN (Table 3.12), LVQ1 (Table 3.13) and LVQ2 (3.14).

   The best accuracy result for the ANN classifiers is 83% with the LVQ2 in table 3.14, but this result is not far from the results of LVQ1 and PNN. Which is a very encouraging result, given that we have not removed "difficult" subjects from the data collection. However, this result falls behind the Diverse Adaboost reported in table 3.9.

   Regarding the usefulness of the features extracted, it is difficult to make an assessment, because some algorithms work better with VV than with MSD, and other have the inverse performance. Training and validation on MSD features is obviously more time efficient and also the best result corresponds to this feature extraction process.

Feature extracted	Features	Hidden units	Accuracy	Sensitivity	Specificity
MSD	24	10	78	0.69	0.88
VV	3611	10	78	0.72	0.84

Table 3.10: Classification results with a BP network with resilient backpropagation.

Feature extracted	Features	Spread	Accuracy	Sensitivity	Specificity
MSD	24	0.02	66%	0.65	0.68
VV	3611	0.852	72.5%	0.65	0.80

Table 3.11: Classification results with a RBF network.

Feature extracted	Features	Spread	Accuracy	Sensitivity	Specificity
MSD	24	0.02	77.8%	0.62	0.94
VV	3611	0.852	74.2%	0.68	0.81

Table 3.12: Classification results with a PNN network.

Feature extracted	Features	Hidden units	Accuracy	Sensitivity	Specificity
MSD	24	10	81%	0.72	0.90
VV	3611	10	79.3%	0.76	0.82

Table 3.13: Classification results with a LVQ1 network. Network training parameters: MSD: 200 epochs, goal: 0.01 and learning rate: 0.01; VV: 150 epochs, goal: 0.10 and learning rate: 0.010.

Feature extracted	Features	Hidden units	Accuracy	Sensitivity	Specificity
MSD	24	10	83%	0.74	0.92
VV	3611	10	77%	0.76	0.78

Table 3.14: Classification results with a LVQ2 network. Network training parameters: MSD: 200 epochs, goal: 0.01 and learning rate: 0.01; VV: 50 epochs, goal: 0.01 and learning rate: 0.005.

3.8   Conclusions and further work

In this section we have studied feature extraction processes based on VBM analysis, to classify MRI volumes of AD patients and normal subjects. We have analyzed different designs for the SPM of the VBM and we have found that the basic GLM design without covariates can detect subtle changes between AD patients and controls that lead to the construction of simple (SVMs and ANNs) and meta-classifiers (cluster AdaBoost and diverse AdaBoost).

Using simple classifiers we have obtained a discriminative accuracy of 83% (table 3.14), in the best case, with a LVQ2 network. Removing difficult subjects with very mild AD from the AD patients subset, this accuracy improves up to 87.5% (table 3.6) with a nonlinear SVM.

In the case of meta-classifiers the accuracy is 86% (table 3.8) if we consider a combination of individual weighted SVM classifiers, with nonlinear kernels. The weighted cluster SVM and the Diverse AdaBoostSVM methods improved remarkably the results, mainly the sensitivity of the classification models.

These accuracy results are really encouraging considering the number of subjects in the database. We have also found that the subjects wrongly classified maybe the most critical ones: old control subjects classified as AD (FP) and subjects with a very early or mild dementia classified as normal (FN), exactly the ones which are the target in these studies that try to perform early detection of AD.

In [43] they compare their results on a smaller population of controls and AD patients to the ones obtained with a standard VBM analysis, using a cluster and found a classification accuracy of 63.3% via cross-validation. Therefore, the results shown in this chapter, along with the careful experimental methodology employed, can be of interest for the Neuroscience community researching on the AD diagnosis based on MRI.

Further work may address the extraction of features based on other morphological measurement techniques, such as the Deformation-Based Morphometry (DBM) and Tensor-Based Morphometry (TBM).

Chapter 4

Lattice Computing for fMRI analysis

In this chapter we introduce a Lattice Computing based approach to the analysis of fMRI. It corresponds to our more recent works. Contrary to previous chapters, the PhD Thesis work described is this chapter is open, showing a broad avenue for future research, and, thus, the results must be taken as a proof of concept rather than an exhaustive exploration.

Section 4.1 gives some background motivation and state of the art. Section 4.2 gives the summary description of the approach tested. Section 4.3 reviews the theoretical foundations and the original endmember induction algorithm. Section 4.4 reports a computational experiment on a case study that illustrate our proposal.

4.1   Background motivation and state of the art

Human brain mapping is a rapidly expanding discipline, and in recent years interest has grown in novel methods for imaging human brain functionality. Non-invasive techniques can measure cerebral physiologic responses during neural activation. One of the relevant techniques is functional Magnetic Resonance Imaging (fMRI) [122], which uses the blood oxygenation level dependent (BOLD) contrast. Slight physiological alterations, such as neuronal activation resulting in changes of blood flow and blood oxygenation, are detected. These signal changes are related to changes in the concentration of deoxyhemoglobin, which acts as an intravascular contrast agent for fMRI. Most of the fMRI examinations are performed with BOLD-based methods using techniques sensitive to local distortions

in the magnetic field (susceptibility sensitive techniques). These are T2 weighted spin echo pulse sequences or T2* weighted gradient echo pulse sequences. The various fMRI-methods have a good spatial and temporal resolution, limited only by the precision with which the autoregulatory mechanisms of the brain adjust blood flow in space to the metabolic demands of neuronal activity. Since these methods are completely noninvasive, using no contrast agent or ionizing radiation, repeated single-subject studies are becoming feasible [117].

To evaluate the resulting fMRI image series, sophisticated algorithms and great computational power are needed to separate the physiologically induced signals from noise or from artifacts resulting from patient movement or MRI detection techniques [45, 118, 108, 143, 150]. Appropriate postprocessing procedures for fMRI are currently being developed at a very rapid pace. Since many research groups are working in this area, no consensus has been reached about the analysis methods of the functional data up to now. A further reason for the large variety of different postprocessing procedures is the lack of a complete underlying theory of the BOLD effect.

The fMRI experiment consists of a functional template or protocol (e.g., alternating activation and rest for a certain time) that induces a functional response in the brain. The aim of an fMRI experiment is to detect this stimulus response, resulting from the BOLD effect, in a defined volume element (voxel). The two main experimental designs used in fMRI are capable of effectively addressing two different questions of interest. Block designs are especially useful for detection, that is, locating which voxels are activated in response to a given task, compared to a control condition. Event-related designs, by contrast, provide a means of estimating the hemodynamic response function. This, in turn, can also lead to detection, as in (for just one example) [71]. The functional information of a voxel has to be extracted from its functional time course. Therefore, for each functional time point one fMRI volume is recorded. The complete four-dimensional dataset (three dimensions in space, one dimension in time) consists of subsequently recorded three-dimensional (3-D) volumes and thus for each voxel of a volume a functional time course exists. The acquisition of these functional volumes runs over periods lasting up to several minutes.

There are a number of sources of noise in the fMRI signal [156] that must be dealt with in appropriate preprocessing steps. The pulse sequence and the magnetic field strength used can have an effect on the image quality. The long time duration of the experiments allow for head motions, even with strong head restraints put into place. Experiment designs also affect the relative dynamics. We will assume that these noise sources have been dealt with appropriately.

The most extended analysis approach for fMRI signals is the Statistical Parametric Map (SPM) [63] which has been implemented in several free open source software packages. This method consists in the separate voxel based test of the general linear model (GLM) corresponding to the experimental design, followed by a segmentation of the spatial distribution of the individual voxel t-test values as a parametric map. There have been also approaches to the fMRI analysis based on the Independent Component Analysis (ICA) [22, 87, 89][30, 35, 34, 33, 80] assuming that the time series observations are linear mixtures of independent sources which can not be observed. ICA assumes that the source signals are non-Gaussian and that the linear mixing process is unknown. The approaches to solve the ICA process obtain both the independent sources and the linear unmixing matrix.

The proposal contained in this chapter is a two step process that consists first in the induction of the endmembers from the fMRI data followed by the linear unmixing of the fMRI data based on the induced endmembers. For the endmembers induction we apply the Endmember Induction Heuristic Algorithm (EIHA) [79]. The basic assumption in this approach is that the data is generated as a convex combination of a set of endmembers which are the vertices of a convex polytope covering the data observations. This assumption is similar to the linear mixture assumed by the ICA or the GLM approaches. The EIHA is based on the relation between the Lattice Independence and Affine Independence [134], and the ability of Lattice Associative Memories to serve as detectors of Lattice Independent sets of vectors. The original works were devoted to unsupervised hyperspectral image segmentation, therefore the use of the name endmember for the convex set vertices. The results obtained on a well known benchmark fMRI data set show the feasibility and promise of this approach. This approach falls in the field of Lattice Computing algorithms, which have been introduced in [77] as the class of algorithms that either apply lattice operators inf and sup or use lattice theory to produce generalizations or fusions of previous approaches.

4.2 Description of the approach

The proposed approach is composed of two steps, as illustrated in algorithm 4.1. For ease of notation we assume that the fMRI data is organized as a matrix whose rows are voxel's time series, after spatial registration and convenient preprocessing. The algorithm EIHA is applied to induce the endmembers from the data. Those endmembers define a convex polytope covering much of the data sam-

ple. After that, the unmixing process obtains the abundance coefficients of each endmember at each voxel by Unconstrained Least Squares Estimation (ULSE). Finally, to give the detection of significant voxels we set a significance threshold for each endmember as the 99% percentil value of the empirical distribision of the abundance coefficients for this endmember.

Algorithm 4.1 Lattice Independent Component Analysis

Given a fMRI data organized as a set of time series $X \in \mathbb{R}^{N \times T}$, where N is the number of voxels and T the time duration

1. Apply EIHA to obtain endmembers $E = \mathbb{R}^{c \times T}$

2. For each voxel compute the endmember abundance coefficients by ULSE, obtaining $A = \mathbb{R}^{N \times c}$.

3. For each abundance volume $A\left(.,k\right) = \mathbb{R}^{N}$ detect the statistical significant voxels as follows:

    (a) Compute the empirical distribution of the abundance values
    (b) Set the significance threshold to the 99% percentil value.

4.3 Theoretical foundations

4.3.1 The linear mixing model

The linear mixing model (LMM) [91, 92] can be expressed as follows:

$$\mathbf{x} = \sum_{i=1}^{M} a_i \mathbf{s}_i + \mathbf{w} = \mathbf{S}\mathbf{a} + \mathbf{w}, \tag{4.1}$$

where $\mathbf{x}$ is the d-dimension pattern vector, $\mathbf{S}$ is the $d \times M$ matrix whose columns are the d-dimension vertices of the convex region covering the data corresponding to the so called endmembers $\mathbf{s}_i, i = 1, .., M$, $\mathbf{a}$ is the M-dimension abundance vector, and $\mathbf{w}$ is the d-dimension additive observation noise vector.

The LMM is applied when some item is assumed to be the combination of several pure items, called endmembers. In [91, 92] the items are light spectra in the

context of hyperspectral image processing, here the items are the fMRI voxel time series vectors. Abundance coefficients correspond to the fraction of the contribution of each endmember to the observed item. From this physical interpretation, follows that the linear mixing model is subjected to two constraints on the abundance coefficients. First, to be physically meaningful, all abundance coefficients must be non-negative $a_i \geq 0, i = 1, .., M$, because the negative contribution is not possible in the physical sense. Second, to account for the entire composition, they must be fully additive $\sum_{i=1}^{M} a_i = 1$. As a side effect, there is a saturation condition $a_i \leq 1, i = 1, .., M$, because no isolate item can account for more than the existent material. From a geometrical point of view, these restrictions mean that we expect the endmembers in $\mathbf{S}$ to be affinely independent and that the convex region defined by them covers *all* the data points. This is a very important observation, because it has deep implications in the following reasoning about the inversion processes. The mixing inversion process (often called unmixing) consists in the estimation of the abundance coefficients, given the endmembers $\mathbf{S}$ and the observation data $\mathbf{x}$. The simplest approach is the unconstrained least squared error (ULSE) estimation given by:

$$\widehat{\mathbf{a}} = \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T\mathbf{x}. \tag{4.2}$$

The coefficients that result from equation (4.2) do not necessarily fulfill the non-negativity and full additivity conditions. The full additivity restriction can be incorporated in the abundance coefficients estimation using Lagrange multipliers [91, 92] introducing a correction term that moves the ULSE estimation to the hyperplane that satisfies the full additivity constraint. From the physical interpretation point of view, the non-negativity restriction is more fundamental. The Non-Negative Least Square estimation (NNLS) [102] can be used to enforce this condition. The estimation problem is treated as a quadratic programming problem with linear inequalities as constraints, solved iteratively. In each iteration the endmembers whose abundances are positive are used to refine the estimation. It was shown in [91, 92] that non-negative estimations do not fulfill the full additivity condition. Besides, the NNLS computation time may be high, specially as the data dimension grows.

Some approaches to endmember determination (i.e., [134]) ensure that the computed endmembers define a convex polytope that covers *all* the data points, so that a proper convex inversion can be attempted. They find that NNLS provide meaningful abundances. However, they also found that the full additivity was not fulfilled. The heuristic algorithm EIHA described in section 4.3.3 [79] always

produces convex regions that lie *inside* the data cloud, so that enforcing the non-negative and full additivity restrictions would be impossible for some data points. Enforcing them for some points may introduce undesired distortions of their abundance values. Nevertheless we have made some attempts to apply NNLS to our data which have resulted in prohibitive computational times.

For the above reasons, we use systematically the unconstrained estimation of equation (4.2) to compute the abundance coefficients. For visualization and interpretation of results, negative abundance values are considered as zero values. We assume that the additivity to one condition is not important as long as we are trying to define statistical tests on the significance of the coefficients based on their distribution as a kind of statistical map. Therefore, abundance coefficients are interpreted as fMRI voxel activation. That is, high positive values above a set threshold are interpreted as high voxel activation of the corresponding endmember time response pattern.

4.3.2 Lattice Independence and Lattice Autoassociative Memories

The work on Lattice Associative Memories (LAM) stems from the consideration of the algebraic lattice structure $(\mathbb{R}, \vee, \wedge, +)$ as the alternative to the algebraic framework given by the mathematical field $(\mathbb{R}, +, \cdot)$ for the definition of Neural Networks computation. The LAM were first introduced in [132, 130] as Morphological Associative Memories, but we follow the new convention introduced in [131, 134] because it sets the works in the more general framework of Lattice Computing. The operators $\vee$ and $\wedge$ denote, respectively, the discrete $\max$ and $\min$ operators (resp. $\sup$ and $\inf$ in a continuous setting). Given a set of input/output pairs of pattern $(X, Y) = \left\{ \left(\mathbf{x}^\xi, \mathbf{y}^\xi \right) ; \xi = 1, .., k \right\}$, a linear heteroassociative neural network based on the pattern's cross correlation is built up as $W = \sum_\xi \mathbf{y}^\xi \cdot \left(\mathbf{x}^\xi \right)'$. Mimicking this constructive procedure [132, 130] propose the following constructions of Lattice Memories (LM):

$$W_{XY} = \bigwedge_{\xi=1}^{k} \left[\mathbf{y}^\xi \times \left(-\mathbf{x}^\xi \right)' \right] \text{ and } M_{XY} = \bigvee_{\xi=1}^{k} \left[\mathbf{y}^\xi \times \left(-\mathbf{x}^\xi \right)' \right], \qquad (4.3)$$

where $\times$ is any of the $\boxtimes$ or $\boxtimes$ operators. Here $\boxtimes$ and $\boxtimes$ denote the $\max$ and $\min$ matrix product [132, 130]. respectively defined as follows:

$$C = A \boxvoid B = [c_{ij}] \Leftrightarrow c_{ij} = \bigvee_{k=1,\dots,n} \{a_{ik} + b_{kj}\}, \tag{4.4}$$

$$C = A \boxbackslash B = [c_{ij}] \Leftrightarrow c_{ij} = \bigwedge_{k=1,\dots,n} \{a_{ik} + b_{kj}\}. \tag{4.5}$$

If $X = Y$ then the LM memories are Lattice Autoassociative Memories (LAM). Conditions of perfect recall by the LM and LAM of the stored patterns proved in [132, 130] encouraged the research on them, because in the continuous case, the LAM is able to store and recall any set of patterns: $W_{XX} \boxvoid X = X = M_{XX} \boxbackslash X$, for any X. However, this result holds when we deal with noise-free patterns. Research on robust recall [125, 130, 133] based on the so-called kernel patterns lead to the notion of morphological independence, in the erosive and dilative sense, and finally to the definition of Lattice Independence (LI) and Strong Lattice Independence (SLI). We gather some results from [134] that set the theoretical background for the approach to endmember induction applied.

Definition Given a set of vectors $\{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$ a *linear minimax combination* of vectors from this set is any vector $\mathbf{x} \in \mathbb{R}^n_{\pm\infty}$ which is a *linear minimax sum* of these vectors:

$$x = \mathcal{L}\left(\mathbf{x}^1, \dots, \mathbf{x}^k\right) = \bigvee_{j \in J} \bigwedge_{\xi=1}^{k} \left(a_{\xi j} + \mathbf{x}^\xi\right),$$

where J is a finite set of indices and $a_{\xi j} \in \mathbb{R}_{\pm\infty} \ \forall j \in J$ and $\forall \xi = 1, \dots, k$.

Definition The *linear minimax span* of vectors $\{\mathbf{x}^1, \dots, \mathbf{x}^k\} = X \subset \mathbb{R}^n$ is the set of all linear minimax sums of subsets of X, denoted $LMS\left(\mathbf{x}^1, \dots, \mathbf{x}^k\right)$.

Definition Given a set of vectors $X = \{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$, a vector $\mathbf{x} \in \mathbb{R}^n_{\pm\infty}$ is *lattice dependent* if and only if $\mathbf{x} \in LMS\left(\mathbf{x}^1, \dots, \mathbf{x}^k\right)$. The vector $\mathbf{x}$ is *lattice independent* if and only if it is not lattice dependent on X. The set X is said to be *lattice independent* if and only if $\forall \lambda \in \{1, \dots, k\}$, $\mathbf{x}^\lambda$ is lattice independent of $X \backslash \{\mathbf{x}^\lambda\} = \{\mathbf{x}^\xi \in X : \xi \neq \lambda\}$.

The definition of lattice independence supersedes and improves the early definitions [133] of erosive and dilative morphological independence, which, however, have more intuitive appealing. Nevertheless, this definition has the additional advantage of establishing a formal parallelism with the definition of linear independence.

Definition A set of vectors $X = \left\{ \mathbf{x}^1, ..., \mathbf{x}^k \right\} \subset \mathbb{R}^n$ is said to be *max dominant* if and only if for every $\lambda \in \{1, ..., k\}$ there exists and index $j_\lambda \in \{1, ..., n\}$ such that

$$x_{j_\lambda}^\lambda - x_i^\lambda = \bigvee_{\xi=1}^{k} \left(x_{j_\lambda}^\xi - x_i^\xi \right) \forall i \in \{1, ..., n\}.$$

Similarly, X is said to be *min dominant* if and only if for every $\lambda \in \{1, ..., k\}$ there exists and index $j_\lambda \in \{1, ..., n\}$ such that

$$x_{j_\lambda}^\lambda - x_i^\lambda = \bigwedge_{\xi=1}^{k} \left(x_{j_\lambda}^\xi - x_i^\xi \right) \forall i \in \{1, ..., n\}.$$

The expressions that compound this definition appeared in the early theorems about perfect recall of Morphological Associative Memories [132, 130]. Their value as an identifiable property of the data has been discovered in the context of the formalization of the relationship between strong lattice independence, defined below, and the classical affine independence.

Definition A set of lattice independent vectors $\left\{ \mathbf{x}^1, ..., \mathbf{x}^k \right\} \subset \mathbb{R}^n$ is said to be *strongly lattice independent* (SLI) if and only if X is max dominant or min dominant or both.

As said before, min and max dominance are the conditions for perfect recall. Per construction, the column vectors of Lattice Autoassociative Memories are min or max dominant, depending of their erosive or dilative nature, therefore they will be strongly lattice independent, *if* they are lattice independent.

Conjecture 4.3.1 *[134] If $X = \left\{ \mathbf{x}^1, ..., \mathbf{x}^k \right\} \subset \mathbb{R}^n$ is strongly lattice independent then X is affinely independent.*

This conjecture (stated as theorem in [131]) is the key result whose proof would relate the linear convex analysis and the nonlinear lattice analysis. If true, it means that the construction of the LAM provides the starting point for obtaining sets of affine independent vectors that could be used as endmembers for the unmixing algorithms described in section 4.3.1.

Theorem 4.3.2 *[134] Let $X = \left\{ \mathbf{x}^1, ..., \mathbf{x}^k \right\} \subset \mathbb{R}^n$ and let W (M) be the set of vectors consisting of the columns of the matrix W_{XX} (M_{XX}.). Let $F(X)$ denote the set of fixed points of the LAM constructed from set X. There exist $V \subset W$ and $N \subset M$ such that V and N are strongly lattice independent and $F(X) = F(V) = F(N)$ or, equivalently, $W_{XX} = W_{VV}$ and $M_{XX} = M_{NN}$.*

The key idea of this theorem is to test the lattice independence of the already known as min or max dominant sets of vectors. Removing lattice dependent vectors will not affect this min/max dominance property. The smart way to test lattice dependence lies in the fact that removing a lattice dependent vectors does not alter the set of fixed points of the remaining ones. This theorem is proved following a constructive reasoning, giving way to an algorithm for the construction of the set of affine independent sets of vectors from LAM dicussed in [79, 134].

4.3.3 Endmember Induction Heuristic Algorithm (EIHA)

For the sake of completeness we recall here our Endmember Induction Heuristic Algorithm (EIHA) maintaining the notation used in the original references. Let us denote $\left\{ \mathbf{f}\left(i\right) \in \mathbb{R}^d : i = 1, .., n \right\}$ the high dimensional data that may be the time series in a fMRI voxels, $\overrightarrow{\boldsymbol{\mu}}$ and $\overrightarrow{\boldsymbol{\sigma}}$ are, respectively, the mean vector and the vector of standard deviations computed componentwise over the voxels of data sample, α the noise correction factor, and E the set of already discovered endmembers. The noise amplitude of the additive noise in equation (4.1) is $\overrightarrow{\boldsymbol{\sigma}}$, the patterns are corrected by the addition and subtraction of $\alpha \overrightarrow{\boldsymbol{\sigma}}$, before being presented to the LAM's. The gain parameter α controls the amount of flexibility in the discovering of new endmembers. For $\mathbf{e} \in \mathbb{R}^d$, let $\mathbf{b}\left(\mathbf{e}\right)$ denote the binary version of $\mathbf{e}$ defined by $b\left(\mathbf{e}\right)_i = 1$ if $e_i > 0$ and $b\left(\mathbf{e}\right)_i = 0$ if $x_i \leq 0$, where $b\left(\mathbf{e}\right)_i$ denotes the ith coordinate of $\mathbf{b}\left(\mathbf{e}\right)$. Finally, we will denote by X the set of binary signatures used to build the lattice memories, and I the set of sample indices corresponding to the endmembers selected by the algorithm.

The detailed description of the steps in the heuristic algorithm is presented as Algorithm 4.2. The starting endmember set consists of a randomly picked pixel. However, this selection is not definitive, because the algorithm may later change this endmember for another, more extreme, one. The noise correction parameter α has a great impact on the number of endmembers found. Low values imply large number of endmembers. It determines if a vector is interpreted as a random perturbation of an already selected endmember.

This algorithm does not need *a priori* information about the nature of the data points that we want to detect. It runs once over the image and finds the most salient data samples on the fly. For this reason we say that it is an on-line algorithm.

Algorithm 4.2 Endmember Induction Heuristic Algorithm (EIHA)

1. Shift the data sample to zero mean
   $\left\{\mathbf{f}^c\left(i\right)=\mathbf{f}\left(i\right)-\overrightarrow{\boldsymbol{\mu}}\,;i=1,..,n\right\}$.

2. Initialize the set of endmembers $E=\left\{\mathbf{e}^1=\mathbf{f}^c\left(i^*\right)\right\}$ where i^* is a randomly picked sample index. Initialize the set of lattice independent binary signatures $X=\left\{\mathbf{x}^1\right\}$ where $\mathbf{x}^1=\mathbf{b}\left(\mathbf{e}^1\right)$. The initial set of endmember sample indices is $I=\left\{i^*\right\}$.

3. Construct the LAM's based on the lattice independent binary signatures: M_{XX} and W_{XX}.

4. For each pixel $\mathbf{f}^c\left(i\right)$

   (a) Compute the noise corrections sign vectors$\mathbf{f}^+\left(i\right)=\mathbf{b}\left(\mathbf{f}^c\left(i\right)+\alpha\overrightarrow{\boldsymbol{\sigma}}\right)$ and $\mathbf{f}^-\left(i\right)=\mathbf{b}\left(\mathbf{f}^c\left(i\right)-\alpha\overrightarrow{\boldsymbol{\sigma}}\right)$

   (b) Compute $y^+=M_{XX}\,\boxtimes\,\mathbf{f}^+\left(i\right)$

   (c) Compute $y^-=W_{XX}\,\boxtimes\,\mathbf{f}^-\left(i\right)$

   (d) If $y^+\notin X$ or $y^-\notin X$ then $\mathbf{f}^c\left(i\right)$ is a new endmember to be added to E, execute once 3 with the new E and resume the exploration of the data sample. Add i to the set of indices I.

   (e) If $y^+\in X$, let k be the index in E of the corresponding endmember. If $\mathbf{f}^c\left(i\right)>\mathbf{e}^k$ then execute step 4g.

   (f) If $y^-\in X$, let k be the index in E of the corresponding endmember. If $\mathbf{f}^c\left(i\right)<\mathbf{e}^k$ then execute step 4g.

   (g) The new data sample is more extreme than the stored endmember, then substitute $\mathbf{e}^k$ in E with $\mathbf{f}^c\left(i\right)$. Index i substitutes the corresponding index in I.

5. The output set of endmembers is the set of original data vectors $\left\{\mathbf{f}\left(i\right):i\in I\right\}$ corresponding to the sign vectors selected as members of E.

4.4 Computational results

The experimental data corresponds to auditory stimulation test data of single person. It is freely available from ftp://ftp.fil.ion.ucI.ac.uk/spm/data, the file name is snrfM00223.zip. These data are the result of the preprocessing pipeline that removes many noise sources. These whole brain BOLD/EPI images were acquired on a modified 2T Siemens MAGNETOM Vision system. Each acquisition consisted of 64 contiguous slices. Each slice being a 2D image of one head volume cut. There are 64x64x64 voxels of size 3mm x 3mm x 3mm. The data acquisition took 6.05s, with the scan-to-scan repeat time (RT) set arbitrarily to 7s. 96 acquisitions were made (RT=7s) in blocks of 6, i.e., 16 42s blocks. The condition for successive blocks alternated between rest and auditory stimulation, starting with rest. Auditory stimulation was bi-syllabic words presented binaurally at a rate of 60 per minute. The functional data starts at acquisition 4, image snrfMOO223-004. Due to T1 effects it is advisable to discard the first few scans (there were no "dummy" lead-in scans). We have discarded the first 10 scans. Figure 4.1 shows the plots of the time series corresponding to the slice #30 of the collected volume. It can be appreciated that there are an intensity displacement filling the whole range of intensities. There are few voxels showing an activation pattern on the top of the plots, and the vast majority of the voxels time series correspond to random non activation patterns at diverse mean intensities. The result of our algorithm applied to these raw data would be trivial and uninteresting.

We perform a data magnitude normalization, consisting in subtracting the mean value of its time series to each voxel independently, so that the plots are collapsed as shown in figure 4.2 around the origin. It can be appreciated that most deactivated voxels are collapsed into a quite similar pattern, and that the diverse activation patterns stand out. This mean subtraction corresponds to an scale normalization in the lattice computing sense. It removes scale effects that hinder the detection of meaningful lattice independent vectors. We note that, although we are shifting the voxel vector to the origin, we are not performing a statistical normalization

The application of the EIHA algorithm with $\alpha = 20$ to the lattice normalized time series of the whole 3D volume produces the collection of eleven endmembers shown in figure 4.3. Attending to the intensity scale it can be assumed that the first endmember (top left plot) corresponds to the non activation pattern, while the remaining endmembers correspond to some kind of activation pattern. These patterns correspond to individual voxels and do not reflect aggregated spatial behaviors like in other approaches.
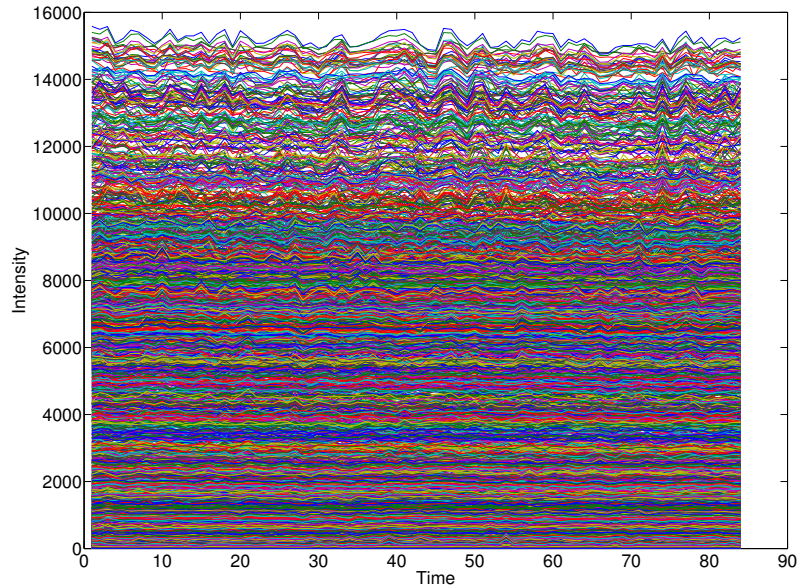
Figure 4.1: Plot of the time series for the voxels of axial slice #30.
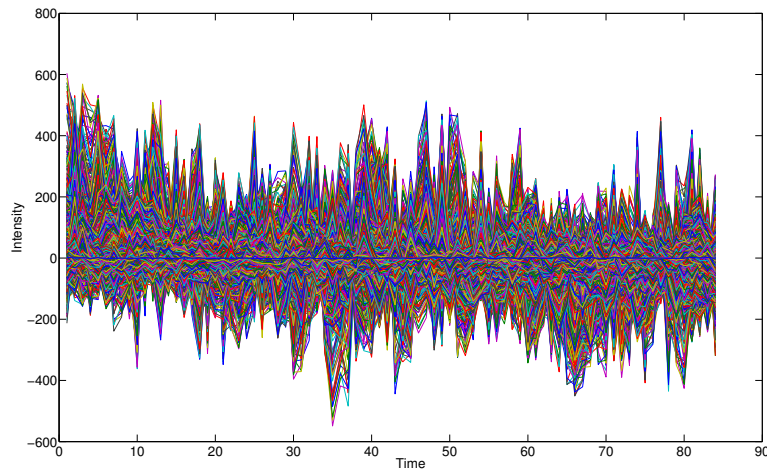
Figure 4.2: Plots of time series of voxels in axial slice #30 after subtracting their mean values from them. The time series are collapsed in the neighborhood of zero.
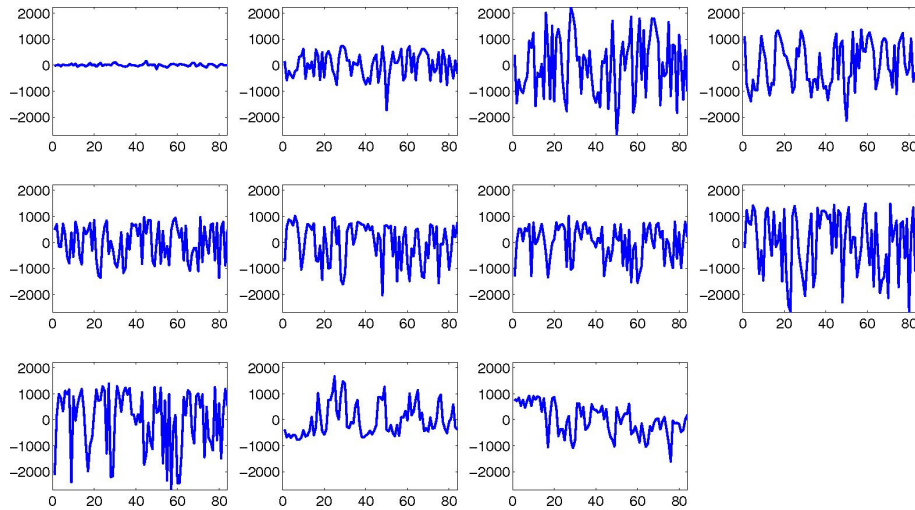
Figure 4.3: Eleven endmbers detected by EIHA over the lattice normalized time series of the whole 3D volume.

The unmixing process applied to the whole volume voxels with the eleven endmembers of figure 4.3 produces the abundance images that we interpret as the activation levels of each pattern. To give an interpretation of these activation levels, we refer to the standard results obtained with the SPM software, presented in figure 4.4 as localized in the Talairach space, in sagital, coronal and axial cuts. There it can be observed that the activation appears around the axial slice #30. For this reason we present the abundances computed on this slice in figure 4.5. The figure presents the original slice where the voxels with abundance value above the 99% percentile of the distribution of this endmember abundance over the whole volume are set to white. It can be appreciated that the abundances for endmembers #8 and #11 have some activation regions overlapping the standard detections in figure 4.4, as well as showing some spurious activation regions. For a complete review of the activation detected by the endmember #11 abundances we show the 99% percentile detection on all the slices in the axial direction in figure 4.6. The figure shows that there are not many spurious detections in slices corresponding to brain regions far away from the activations shown in figure 4.4. So that our results are in agreement with those found with the SPM software. The main difference is that we have not imposed an activation model, but we discovered on-line the activation pattern to perform the detection.
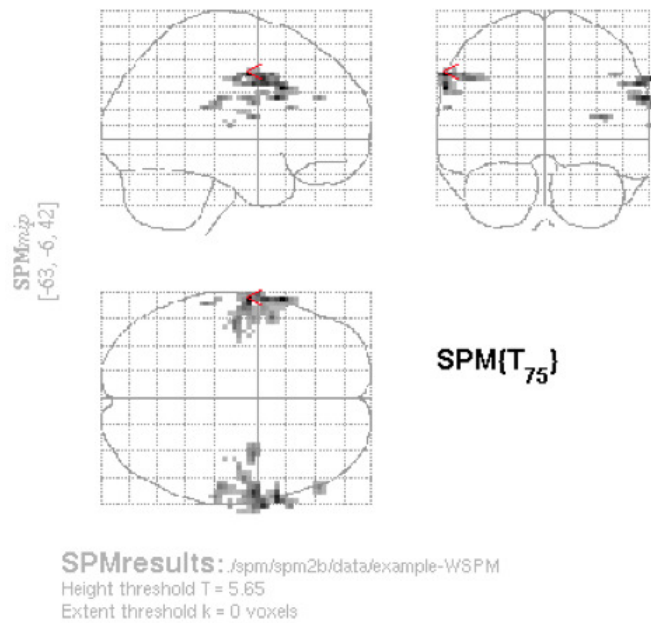
Figure 4.4: Activation maps from SPM results over the experimental data

4.5 Conclusions

In this chapter, we have proposed and applied the endmember induction algorithm EIHA discussed in [79] to the task of on-line brain region activation in fMRI. The idea is inspired in the ICA application to fMRI activation detection [33, 143], in our approach the sources correspond to endmembers detected by the EIHA algorithm and the activation is computed as the abundance images obtained by unmixing the voxel time series on the basis of the found endmembers. The first obstacle that we find in this endeavor is that the distribution of the time series is not well aspected for the detection of lattice independence as a meaningful characteristic. In fact the voxel's fMRI time series show a dense distribution of intensity displacements from the origin, so almost all of them are lattice dependent and our proposed algorithm only recovers two endmembers. To overcome this problem we perform a data magnitude normalization which corresponds to a scale normalization in the sense of lattice computing. We subtract the mean of its time series to each voxel time series. The resulting lattice normalized data set shows a much more rich structure in terms of lattice independence. Our computational experiment with a well known fMRI data set, provided with the distribution of the SPM
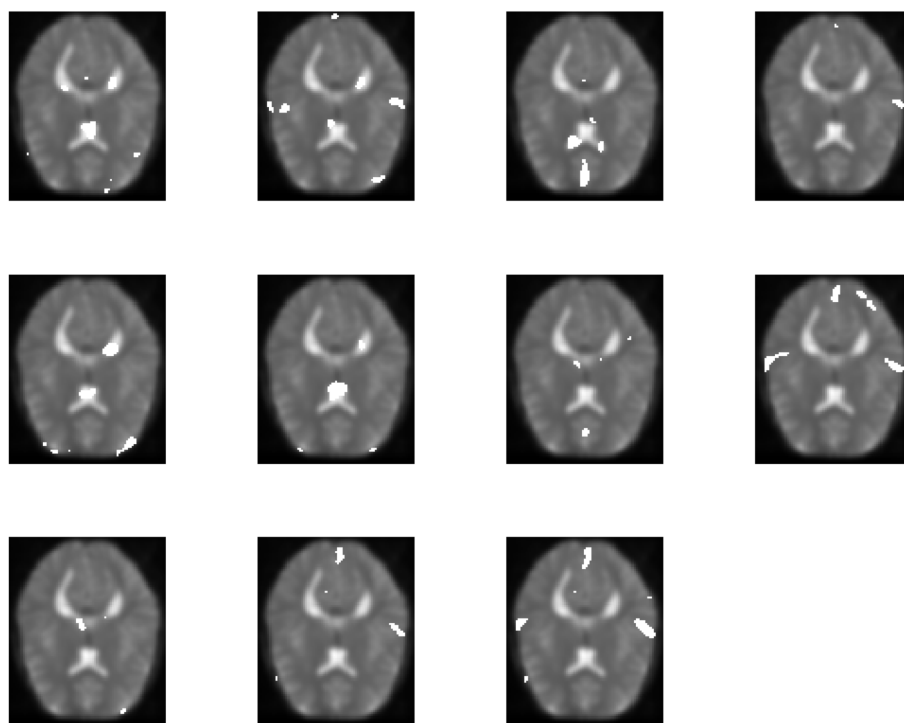
Figure 4.5: Abundances for axial slice #30 for all eleven endmembers. White voxels correspond to abundance values above the 99% percentile of the distribution of the abundances for each endmember at this slice.
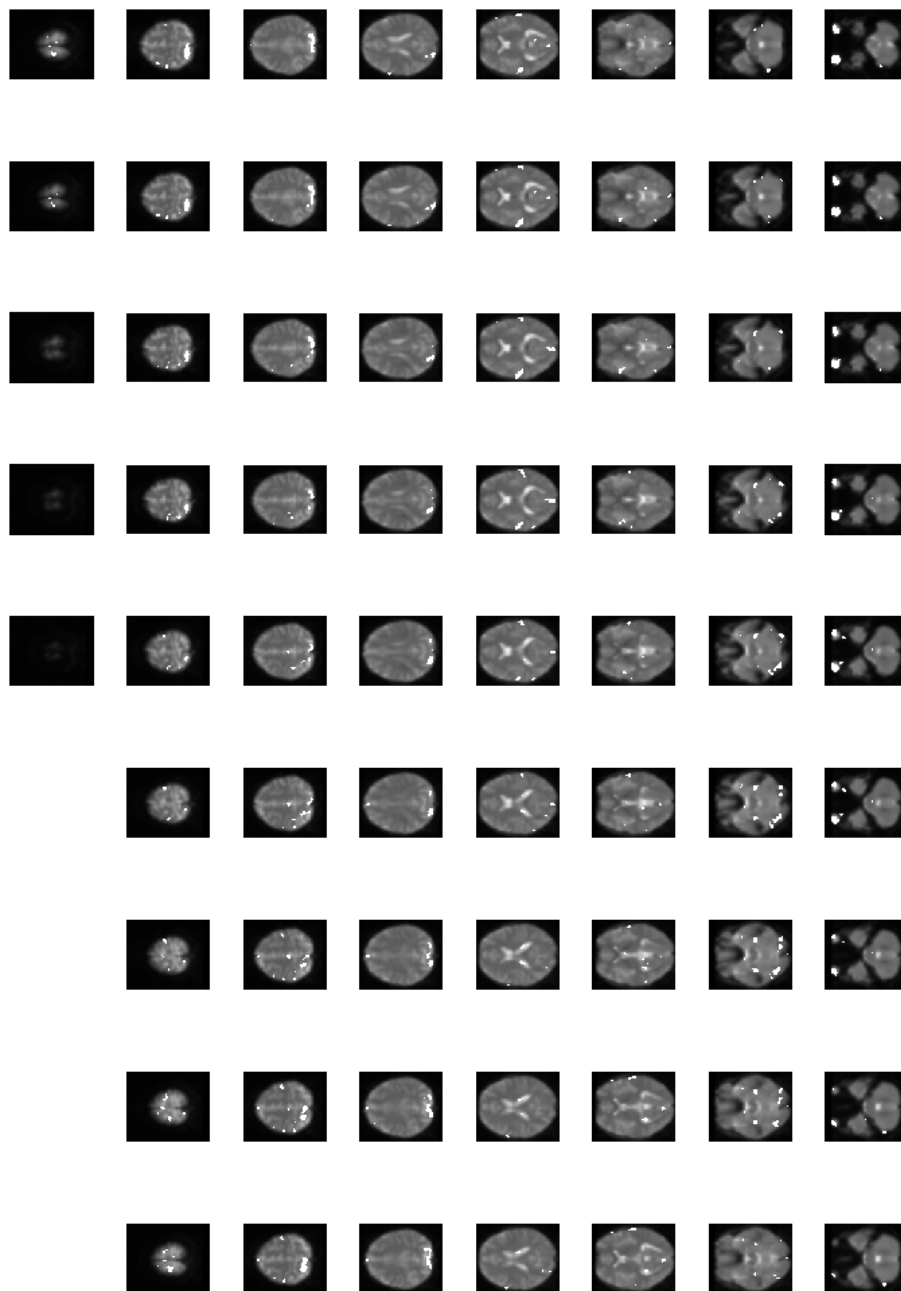
Figure 4.6: Activations detected by the 99% percentile of the abundance images of endmember #11 of figure 4.5 in the axial direction.

software, show some promising results in the sense that our detected activations are in strong agreement with the standard analysis using the SPM software. There are however some false detections that show that our approach is not fully consistent with the SPM analysis. One important aspect of SPM is its process of the individual voxel t-test as a random field, this processing is lacking in our works. Finding ways to harmonize global random field analysis and our lattice computing approach may lead to interpretations and results. This PhD Thesis work will be continued in the framework of the research group. This chapter gives a proof of concept of an approach that, when fully explored, may be as fruitful as the ICA applications to fMRI.

Appendix A

Description of IIH correction algorithms

In this appendix we provide the formal definitions of the IIH correction algorithms that have been implemented to compare results with our own propositions. We maintain the notation of chapter 2, adapting the expressions given in the original papers to this notation. Section A.1 presents the Biased MAP estimation. Section A.2 presents the algorithm proposed by Wells et al. Finally section A.3 presents the Biased FCM.

A.1   BMAP

The Biased MAP (BMAP) [127] algorithm is a bayesian image processing algorithm [67, 68] which consists of two interdependent estimations:

- the Maximum A Posteriori (MAP) estimation of the image classification given an inhomogeneity field estimation, and

- the Maximum Likelihood (ML) estimation of the inhomogeneity field given an image classification.

The algorithm iterations interleave both estimations, performing first the MAP estimation and second the ML estimation at each iteration.

The MAP estimation of the classification image $\mathbf{x}^* = \arg\max_x p\left(\mathbf{x} \,|\, \mathbf{y}\right)$ can be stated as a minimization of the following energy function

$$u\left(\mathbf{x} \,|\, \mathbf{y}\right) \propto -\log\left(p\left(\mathbf{x} \,|\, \mathbf{y}\right)\right),$$

that is:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} u\left(\mathbf{y} \,|\, \mathbf{x}, \beta^*, \theta^*, \alpha^*\right), \tag{A.1}$$

where β^*, θ^* and α^* are, respectively, the estimated inhomogeneity field, the collection of parameters of the Gaussian Conditional Probability Density Functions (GCPDF) and the parameters of the Markov Random Field (MRF). The GCPDF model the likelihood of the observed image under the assumption of the knowledge of the classification image and the IIH field. The MRF provides the model of the *a priori* distribution, embodying some spatial smoothness constraints of the classification image. As proposed in [127], the energy minimization is performed by the application of Iterated Conditional Modes (ICM), which is a greedy local search algorithm. ICM assumes the spatial independence of the voxel intensity distributions defining the following local energy at each voxel:

$$u\left(x_i \,|\, y_i\right) = -\frac{1}{2}\left(\frac{y_i - \left(1 + \beta_i\right)\mu_{x_i}}{\sigma_{x_i}}\right)^2 + \log\left(\sigma_{x_i}\right) + V_1\left(x_i\right) + V_2\left(x_i\right), \tag{A.2}$$

where μ_{x_i} and σ_{x_i} are the mean and standard deviation of the intensity Gaussian distribution associated with class i; V_1 and V_2 are the first and second order potentials of the Markov Random Field that models the *a priori* image distribution. ICM performs the independent minimization of the local energy at each voxel, discarding spatial interaction effects:

$$\min_{x_i} u\left(x_i \,|\, y_i\right). \tag{A.3}$$

The algorithm A.1 provides a pseudo-code specification of ICM.

The complete BMAP algorithm iterates the MAP estimation of the classification $\mathbf{x}(t)$, done by the ICM, and the maximum likelihood estimation of the IIH field, the class intensity averages and their variances, given by the following equations:

$$\widehat{\beta_i} = \frac{y_i}{\mu_{x_i}} - 1, \tag{A.4}$$

$$\widehat{\mu}_\omega = \frac{\sum_{i \in R_\omega} y_i\left(1 + \beta_i\right)}{\sum_{i \in R_\omega}\left(1 + \beta_i\right)^2}, \tag{A.5}$$

$$\widehat{\sigma}_\omega^2 = \frac{\sum_{i \in R_\omega}\left(y_i - \left(1 + \beta_i\right)\mu_\omega\right)^2}{|R_\omega|}. \tag{A.6}$$

Where R_ω denotes the image regions whose voxels are classified as of tissue ω. We omit the iteration index t in these equations for the sake of clarity. Algorithm A.2 provides a pseudocode specification of the BMAP.

Algorithm A.1 ICM algorithm for MAP estimation of $\mathbf{x}(t)$

Given $\hat{\mathbf{x}}(0) \equiv \mathbf{x}(t-1)$
$k = 0$
Repeat

    1. $k = k + 1$

    2. For each i perform minimization specified by equation (A.3)

        (a) $\hat{x}_i(k) = \arg\min_{x_i} u(x_i | y_i)$

Until there is no variation in $\hat{\mathbf{x}}(k)$
Return $\hat{\mathbf{x}}(k)$ as $\mathbf{x}(t)$

Algorithm A.2 BMAP algorithm

Given initial values $\mathbf{x}(0), \widehat{\beta}_i(0), \widehat{\mu}_\omega(0), \widehat{\sigma}^2_\omega(0)$
$t = 0$
Repeat

    1. $t = t + 1$

    2. Application of ICM to approximate the MAP estimation of $\mathbf{x}(t)$

    3. Apply equations (A.4), (A.5) and (A.6) to estimate the bias and Gaussian densities parameters with maximum likelihood

Until $t > T$ or $\mathbf{x}(t) = \mathbf{x}(t-1)$

A.2   Wells

The algorithm proposed by Wells et al. [168] performs the MAP estimation of the bias field (the logarithm of the multiplicative inhomogeneity field):

$$\widehat{\mathbf{B}} = \arg\max_B p\left(\mathbf{B}\,|\,\mathbf{Y}\right). \tag{A.7}$$

The Wells algorithm assumes a given assignment of voxels to tissue classes. The distribution of the intensity values is Gaussian, with known mean μ_ω and variance σ_ω^2 for each tissue class ω. From classical calculus, a necessary condition for a maximum of the *a posteriori* probability of $\mathbf{B}$ is that its gradient with respect to $\mathbf{B}$ is zero. After some formal derivations, this condition leads to the following equality:

$$\widehat{\mathbf{B}} = H\overline{R}. \tag{A.8}$$

The weighted residuals $\overline{R}$ are given by:

$$\left[\overline{R}\right]_i \equiv \sum_\omega W_{i\omega}\sigma_\omega^{-1}\left(Y_i - \mu_\omega\right), \tag{A.9}$$

where $W_{i\omega}$ is equivalent to the *a posteriori* probability of class ω in voxel site i. The factor H is an expression that depends on the covariance matrix of the bias field, and it is in general intractable. The bias field estimator is assumed to be well approximated by

$$\widehat{\mathbf{B}}_i = \frac{\left[F\overline{R}\right]_i}{\left[F\boldsymbol{\sigma}\right]_i'}, \tag{A.10}$$

where F is a low pass filter, σ is the matrix of voxel variances with $\sigma_i \equiv \sum_\omega W_{i\omega}\sigma_\omega^{-1}$. The Wells algorithm is an EM algorithm consisting of the iteration until convergence of the following steps:

$$W_{i\omega} \leftarrow \frac{p\left(\omega\right)G_{\sigma_\omega}\left(Y_i - \mu_\omega - B_i\right)}{\sum_{j=1}^c p\left(\omega_j\right)G_{\sigma_{\omega_j}}\left(Y_i - \mu_{\omega_j} - B_i\right)}. \tag{A.11}$$

$$\widehat{\mathbf{B}} \leftarrow H\overline{R}, \tag{A.12}$$

where $G_\sigma\left(\mu\right)$ denotes the gaussian probability density function as in [168].

A.3 BFCM

The Bias Fuzzy C-Means (BFCM) method was presented in [18, 124, 123]. It is representative of the fuzzy clustering MRI segmentation methods. The method assumes that the image has been logarithmically transformed. The algorithm is designed as the realization of the following minimization process:

$$\min_{U,\{\mu_\omega\}_{\omega\in\Omega},\{B_i\}_{i\in I}} J_m \text{ subject to } U \in \mathcal{U}, \tag{A.13}$$

where $\mathcal{U}$ is the set of consistent partitions, such that $u_{i\omega} > 0$ and $\sum_\omega u_{i\omega} = 1$, and the objective function is defined as follows:

$$J_m = \sum_{\omega\in\Omega}\sum_{i\in I} u_{i\omega}^p \|R_i - \mu_\omega\|^2 + \frac{\alpha}{N_R}\sum_{\omega\in\Omega}\sum_{i\in I} u_{i\omega}^p \left(\sum_{l\in\mathcal{N}_i} \|R_l - \mu_\omega\|^2\right), \tag{A.14}$$

which is an adaptation of the conventional Fuzzy C-Means (FCM) objective function [23, 24, 50]. In this expression, $\{\mu_\omega\}_{\omega\in\Omega}$ are the cluster representatives, and the matrix $[u_{i\omega}] = U$ represents the partition matrix, $U \in \mathcal{U}$. $\mathcal{N}_i$ denotes the voxel i neighborhood of cardinality N_R. Finally, α is a weight given to the regularization term. The zero gradient condition of the cost function of equation (A.14) plus the Lagrange multiplier corresponding to the restrictions on the partition matrix lead to the following expressions of their estimates:

$$\widehat{u}_{i\omega}^* = \frac{1}{\sum_{\theta\in\Omega} \left(\frac{D_{i\omega} + \frac{\alpha}{N_R}\gamma_{i\omega}}{D_{i\theta} + \frac{\alpha}{N_R}\gamma_{i\theta}}\right)^{1/(p-1)}}. \tag{A.15}$$

$$\widehat{\mu}_\omega^* = \frac{\sum_{i\in I} u_{i\omega}^p \left((Y_i - B_i) + \frac{\alpha}{N_R}\left(\sum_{l\in\mathcal{N}_i}(Y_l - B_l)\right)\right)}{(1+\alpha)\sum_{i\in I} u_{i\omega}^p}. \tag{A.16}$$

$$\widehat{B}_i^* = Y_i - \frac{\sum_{i\in I} u_{i\omega}^p \mu_\omega}{\sum_{i\in I} u_{i\omega}^p}. \tag{A.17}$$

In these expressions $D_{i\omega} = \|Y_i - B_i - \mu_\omega\|^2$ and $\gamma_{i\omega} = \sum_{l\in\mathcal{N}_i} \|Y_l - B_l - \mu_\omega\|^2$. In this formulation, intensity class prototypes are a kind of intensity average of the IIH corrected voxels, computed over the neighborhood of each voxel. The bias field is estimated as the residual of the prediction of the observed voxel Y_i based on the intensity class representatives and the membership coefficients. The BFCM algorithm consists in the iteration of the application of the equations (A.15), (A.16)

and (A.17) until convergence in the estimation of the intensity class representatives. The algorithm A.3 gives a pseudo-code specification of the algoritm.

There were no bias smoothing steps explicit in the description of the algorithm [18], however the later modification proposed in [148] includes an heuristic smoothing over the bias estimate which ensures that the spatial gradient of the bias estimate is below a predefined threshold.

Algorithm A.3 BFCM algorithm

Perform the logarithm transformation of the input volume.
Given initial values $U(0)$, $\mathbf{B}(0)$
Repeat

1. $t = t + 1$

2. Compute the estimation of the Fuzzy C-Means applying equation (A.16)

3. Compute the updated membership functions $U(t)$ applying equation (A.15)

4. Compute the updated bias field estimation $\mathbf{B}(t)$ applying equation (A.17)

Until $U(t) = U(t-1)$, and $\mathbf{B}(t) = \mathbf{B}(t-1)$

Appendix B

Machine Learning Methods

In this appendix we will gather formal definitions of the Machine Learning methods employed to build the classification systems tested in chapter 3 on the detection of Alzheimer Disease (AD) in structural MRI.

B.1   Support Vector Machines

The Support Vector Machine (SVM) [163] algorithm used for this study is included in the libSVM (`http://www.csie.ntu.edu.tw/~cjlin/libsvm/`) software package. The implementation is described in detail in [36]. Given training vectors $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, \ldots, l$ of the subject features of the two classes, and a vector $\mathbf{y} \in \mathbb{R}^l$ such that $y_i \in \{-1, 1\}$ labels each subject with its class, in our case, for example, patients were labeled as -1 and control subject as 1. To construct a classifier, the SVM algorithm solves the following optimization problem:

$$\min_{\mathbf{w}, b, \boldsymbol{\xi}} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{l} \xi_i,$$

subject to

$$y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq (1 - \xi_i),\ \xi_i \geq 0,\ i = 1, 2, \ldots, n.$$

The dual optimization problem is:

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha,$$

subject to

$$\mathbf{y}^T \alpha = 0, \ 0 \leq \alpha_i \leq C, \ i = 1, \ldots, l.$$

Where e is the vector of all ones, $C > 0$ is the upper bound on the error, Q is an l by l positive semidefinite matrix:

$$Q_{ij} \equiv y_i y_j K(\mathbf{x}_i, \mathbf{x}_j),$$

and

$$K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j),$$

is the kernel function that describes the behavior of the support vectors. Here, the training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function $\phi(\mathbf{x}_i)$. The decision function is:

$$sgn(\sum_{i=1}^{l} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b).$$

The regularization parameter C is used to balance the model complexity and the training error. It was always set to 1.

The chosen kernel function results in different kinds of SVM with different performance levels, and the choice of the appropriate kernel for a specific application is a difficult task. In this study two different kernels were tested: the linear and the radial basis function (RBF) kernel.

- The linear kernel function is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = 1 + \mathbf{x}_i^T \mathbf{x}_j,$$

this kernel shows good performance for linearly separable data.

- The RBF kernel is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}\right).$$

This kernel is basically suited best to deal with data that have a class-conditional probability distribution function approaching the Gaussian distribution [31]. One of the advantages of the RBF kernel is that given a kernel, the number of support vectors and the support vectors are all automatically obtained as part of the training procedure, i.e., they do not need to be specified by the training mechanism.

B.2 Artificial Neural Networks

B.2.1 Backpropagation

Many researchers worked in the area of adaptive systems during the 1960's using perceptrons and single layer linear neural networks. The single-node perceptron learning rule has been shown to converge when a linearly separable solution to the discrimination problem exists (see pp. 82-87 of [120]). However, in 1969, Minsky and Papert showed that a single node perceptron could not perform the simple Boolean function exclusive-OR [115] discouraging many researchers from further work in the area. From that time until the early 1980's, the shift in AI research was to Expert Systems and Artificial Neural Network research was restricted to the realms of cybernetics communities and, while some research in adaptive systems continued in the areas of control and signal systems. Pointing out that a simple two-layer network can be easily made to perform the exclusive-OR problem [139] was one psychological step to thaw the interest into Artificial Neural Networks. In fact, early works of Rosenblatt developed some algorithms that could train multilayer networks [136], although lacking a convergence proof. The lack of an effective training rule for multilayer networks has been cited by many researchers as the primary reason for the demise of neural network research in the 1970's.

There are now many training algorithms available for multilayer perceptrons. The architecture of a typical multilayer perceptron is as shown in B.2.1. Some algorithms have been developed for multilayer networks where the nodes have hardlimiter nonlinearitics (e.g., MRII by Widrow [171]). However, the most popular architecture uses sigmoidal nonlinearities on the nodes. The sigmoid is differentiable, which makes possible weight update rules based on the gradient of the error with respect to the weights in the network. The most popular rule for training the weights in a multilayer perceptron is the backpropagation (BP) training algorithm. This technique was popularized by [139] although it was first derived by Werbos in 1974 [169] and rederived by Parker in 1982 [121]. Recently, it has been suggested by White [170] that the stochastic approximation techniques developed by Robbins and Monro [135] in 1951 subsume backpropagation.

Backpropagation is a gradient descent method for training the weights in a multilayer perceptron. A regression or discrimination problems characterized by a set of training vectors X such that for every vector $x \in X$, there is an associated desired output vector $d \in D$, where D is the set of desired outputs associated with the training vectors in X. If D is a discrete set then we have a classification problem, if it is a continuous set then we have a regression problem. Let the

instantaneous error E_p be defined as:

$$E_p = \frac{1}{2} \left(\boldsymbol{d}_p - \mathbf{z}_p\right)^T \left(\boldsymbol{d}_p - \mathbf{z}_p\right) = \frac{1}{2} \sum_{k=1}^{N} \left(d_{k.,p} - z_{k,p}\right)^2, \tag{B.1}$$

where $d_{k,p}$ is the k-th component of the p-th desired output vector $\boldsymbol{d}_p$, and $z_{L,k}\left(\mathbf{x}_p\right)$ is the k-th component of the network output vector z_L when the p-th training exemplar $\mathbf{x}_p$ is input to the multilayer perceptron. L is the number of layers of the multilayer feedforward network. The output of the j-th node in layer l is given by:

$$z_{l,j}\left(\mathbf{x}_p\right) = f\left(\sum_{i=0}^{N_{l-1}} w_{l,j,i} z_{l-1,i}\left(\mathbf{x}_p\right)\right), \tag{B.2}$$

where $z_{l,j}$ is the output of node in layer l, N_l is the number of nodes in layer l, $w_{l,j,i}$ is the weight which connects the i-th node in layer $l-1$ to the jth node in layer l, and $f\left(\cdot\right)$ is the sigmoid nonlinearity, which has a simple derivative:

$$f'\left(\alpha\right) = \frac{df\left(\alpha\right)}{d\alpha} = f\left(\alpha\right)\left(1 - f\left(\alpha\right)\right). \tag{B.3}$$

The convection is that $z_{0,k}\left(\mathbf{x}_p\right) = \mathbf{x}_{p,k}$. Let the total error E_T be defined as follows:

$$E_T = \sum_{p=1}^{P} E_p, \tag{B.4}$$

where P is the cardinality of $\boldsymbol{X}$. Note that E_T is a function of both the training set and the weights in the network. The backpropagation learning rule is defined as follows:

$$\triangle w\left(t\right) = -\eta \frac{\partial E_p}{\partial w} + \alpha \triangle w\left(t-1\right), \tag{B.5}$$

where $0 < \eta < 1$, which is the learning rate, the momentum factor α is also a small positive number, and w represents any single weight in the network. In the above equation, $\triangle w\left(t\right)$ is the change in the weight computed at time t. When the momentum term is used $(\alpha \neq 0)$, the training rule is called the momentum method; otherwise, it is the backpropagation method. The algorithm (B.5 with $\alpha = 0$) is often termed as *instantaneous backpropagation* because it computes the

gradient based on a single training vector. Another variation is *batch backprop-agation*, which computes the weight update using the gradient based on the total error E_T. When training results are presented, instantaneous backpropagation is the method used.

To implement this algorithm we must give an expression for the partial derivative of E_p with respect to each weight in the network. For an arbitrary weight in layer l this can be written using the Chain Rule:

$$\frac{\partial E_p\left(\boldsymbol{w}\right)}{\partial w_{l,j,i}} = \frac{\partial E_p\left(\boldsymbol{w}\right)}{\partial z_{l,j}\left(\mathbf{x}_p\right)} \frac{\partial z_{l,j}\left(\mathbf{x}_p\right)}{\partial w_{l,j,i}}, \tag{B.6}$$

where

$$\begin{aligned} \frac{\partial z_{l,j}\left(\mathbf{x}_p\right)}{\partial w_{l,j,i}} &= \frac{\partial}{\partial w_{l,j,i}} \left[f\left(\sum_{m=0}^{N_{l-1}} w_{l,j,m} z_{l-1,m}\left(\mathbf{x}_p\right) \right) \right] \\ &= f'\left(\sum_{m=0}^{N_{l-1}} w_{l,j,m}\, z_{l-1,m}\left(\mathbf{x}_p\right) \right) \frac{\partial}{\partial w_{l,j,i}} \left[\sum_{m=0}^{N_{l-1}} w_{l,j,m}\, z_{l-1,m}\left(\mathbf{x}_p\right) \right] \\ &= f'\left(\sum_{m=0}^{N_{l-1}} w_{l,j,m}\, z_{l-1,m}\left(\mathbf{x}_p\right) \right) z_{l-1,i}\left(\mathbf{x}_p\right). \end{aligned} \tag{B.7}$$

Substituting from equation (B.3) for the first term, we get:

$$\frac{\partial z_{l,j}}{\partial w_{l,j,i}} = z_{l,j}\left(1 - z_{l,j}\right) z_{l-1,j}. \tag{B.8}$$

With this, equation (B.6) becomes:

$$\frac{\partial E_p\left(\boldsymbol{w}\right)}{\partial w_{l,j,i}} = \frac{\partial E_p\left(\boldsymbol{w}\right)}{\partial z_{l,j}} z_{l,j}\left(1 - z_{l,j}\right) z_{l-1,j}. \tag{B.9}$$

The term $\partial E_p(\boldsymbol{w})/\partial z_{l,j}(\mathbf{x}_p)$ represents the sensitivity of $E_p\left(\boldsymbol{w}\right)$ to the output of node $z_{l,j}\left(\mathbf{x}_p\right)$. The node $z_{l,j}$ propagates its influence on $E_p\left(\boldsymbol{w}\right)$ through all of the nodes in the succeeding layers. Thus, $\partial E_p(\boldsymbol{w})/\partial z_{l,j}(\mathbf{x}_p)$ can be expressed as a function of the sensitivities to nodes in the next highest layer as follows:

$$\frac{\partial E_p(\boldsymbol{w})}{\partial z_{l,j}(\mathbf{x}_p)} = \sum_{m=1}^{N_{l+1}} \frac{\partial E_p(\boldsymbol{w})}{\partial z_{l+1,m}(\mathbf{x}_p)} \frac{\partial z_{l+1,m}(\mathbf{x}_p)}{\partial z_{l,j}(\mathbf{x}_p)} \tag{B.10}$$

$$= \sum_{m=1}^{N_{l+1}} \frac{\partial E_p(\boldsymbol{w})}{\partial z_{l+1,m}(\mathbf{x}_p)} \frac{\partial}{\partial z_{l,j}(\mathbf{x}_p)} \left[f\left(\sum_{q=0}^{N_l} w_{l+1,m,q}\, z_{l,q}(\mathbf{x}_p) \right) \right]$$

$$= \sum_{m=1}^{N_{l+1}} \frac{\partial E_p(\boldsymbol{w})}{\partial z_{l+1,m}(\mathbf{x}_p)} f'\left(\sum_{q=0}^{N_l} w_{l+1,m,q}\, z_{l,m}(\mathbf{x}_p) \right) \frac{\partial}{\partial z_{l,j}(\mathbf{x}_p)} \left[\sum_{q=0}^{N_l} w_{l+1,m,q}\, z_{l,q}(\mathbf{x}_p) \right]$$

$$= \sum_{m=1}^{N_{l+1}} \frac{\partial E_p(\boldsymbol{w})}{\partial z_{l+1,m}(\mathbf{x}_p)} z_{l+1,m}(\mathbf{x}_p) \left(1 - z_{l+1,m}(\mathbf{x}_p)\right) w_{l+1,m,j}.$$

This process can be continued for $\partial E_p(\boldsymbol{w})/\partial z_{l+1,m}$ and so on, until we reach the output layer. At the output layer we reach a *boundary condition* where the sensitivities of the nodes in the last layer are derived from the total squared error for the p-th pattern as:

$$\frac{\partial E_p(\boldsymbol{w})}{\partial z_{L,j}} = z_{L,j}(\boldsymbol{x}_p) - d_j(\boldsymbol{x}_p). \tag{B.11}$$

While the derivation seems to be working its way forward to the output layer, the sensitivity of a node is actually computed from the output layer *backwards*. The expression in equation (B.11) is called the output error, and the corresponding expression for hidden layer nodes in equation (B.10) is often referred to as the hidden layer error, although strictly speaking it does not represent an error because we do not known the desired response for some hidden units and we can not compute this error directly.

B.2.2 Probabilistic Neural Networks

The PNNwas first proposed in [153]. The architecture of a typical PNN is as shown in B.2. The input layer unit does not perform any computation and simply distributes the input to the neurons in the *pattern* layer. On receiving a pattern from the input layer, the neuron $\mathbf{x}_{ij}$ of the pattern layer computes its output:

$$\Phi_{ij}(x) = \frac{1}{(2\pi)^{d/2}\, \sigma^d} \exp\left[-\frac{(x - x_{ij})^T (x - x_{ij})}{2\sigma^2} \right], \tag{B.12}$$
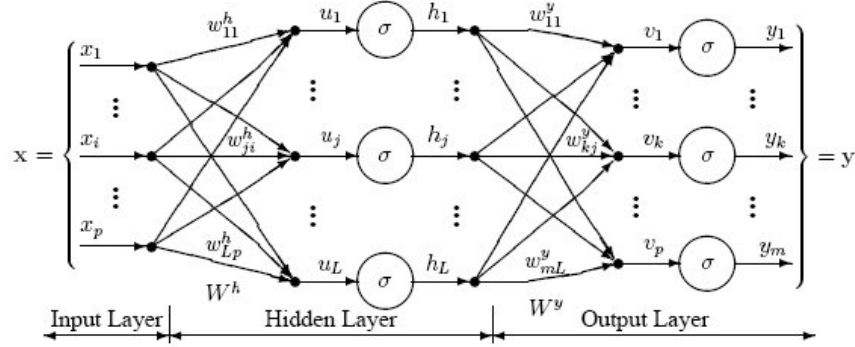
Figure B.1: Architecture of a typical multilayer perceptron.

where d denotes de dimension of the pattern vector x, σ is the smoothing parameter and $\mathbf{x}_{ij}$ is the neuron vector. The output layer neurons compute the likelihood of pattern $\mathbf{x}$ begin classified into C_i by averaging the output of all neurons that belong to the same class:

$$p_i\left(\mathbf{x}\right) = \frac{1}{\left(2\pi\right)^{d/2} \sigma^d} \frac{1}{N_i} \sum_{j=1}^{N_i} \exp\left[-\frac{\left(\mathbf{x}-\mathbf{x}_{ij}\right)^T \left(\mathbf{x}-\mathbf{x}_{ij}\right)}{2\sigma^2}\right], \qquad \text{(B.13)}$$

where N_i denotes the total number of samples in class C_i. If the *a priori* probabilities for each class are the same, and the losses associated with making an incorrect decision for each class are the same, the decision layer unit classifies the pattern x in accordance with the optimal Bayes's decision rule based on the output of all the output layer neurons:

$$\hat{C}\left(x\right) = \arg\max\left\{p_i\left(\mathbf{x}\right)\right\}, \quad i = 1, 2, ..., m. \qquad \text{(B.14)}$$

where $\hat{C}\left(x\right)$ denotes the estimated class of the pattern $\mathbf{x}$ and m is the total number of classes in the training samples. One outstanding issue associated with the PNN is the determination of the network structure. This includes determining the network size, the pattern layer neurons and an appropriate smoothing parameter. Some algorithms for pattern layer neurons selection have been proposed [32, 126, 154, 155, 162, 176]. The conventional approach is to assume that the PNN is a kind of k-NN classifier, with a Gaussian filter imposed upon each data sample pattern. Therefore, all the patterns in the training set are used as pattern

neurons. The Gaussian variance parameter is usually set heuristically because there is no closed form expression relating the classification error.
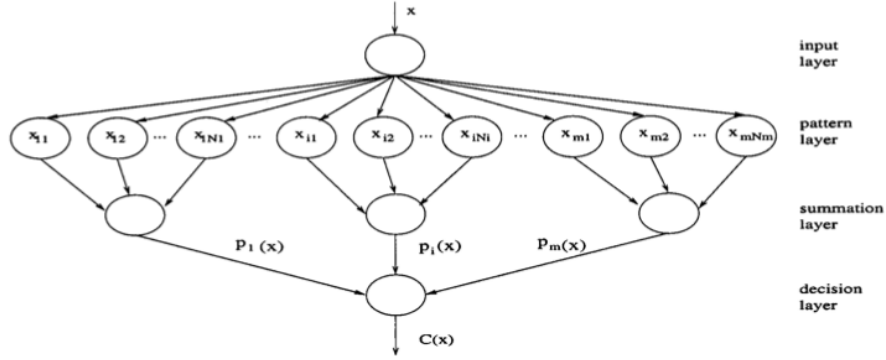
Figure B.2: Architecture of a typical PNN

B.2.3 Radial Basis Functions Network

RBFs have their origin in the solution of the multivariate interpolation problem [27]. Arbitrary function $g(\boldsymbol{v}) : \mathbb{R}^d \to \mathbb{R}$ can be approximated by mapping, using a RBF network with a single hidden layer of p units:

$$\hat{g}(\boldsymbol{v}, \boldsymbol{x}) = \sum_{j=1}^{p} w_j r_j (\boldsymbol{v}, \boldsymbol{\sigma}_j, \boldsymbol{c}_j)$$

$$= \sum_{j=1}^{p} w_j \phi_j (\boldsymbol{\sigma}_j, \|\boldsymbol{v} - \boldsymbol{c}_j\|), \tag{B.15}$$

where $\boldsymbol{v} \in \mathbb{R}^d$; $\boldsymbol{x}$ is the vector of variable factors including w_j, $\boldsymbol{\sigma}_j$, and $\boldsymbol{c}_j$; p denotes the number of basis functions; $\mathbf{w} = (w_1, w_2, ..., w_p)^T$ contains the weight coefficients; $r_j(\cdot)$ represents de d-dimensional activation function (also known as the radial basis function) from $\mathbb{R}^d$ to $\mathbb{R}$; $\|\cdot\|$ is the Euclidean norm; $\boldsymbol{c}_j = (c_{j1}, c_{j2}, \ldots, c_{jd})^T$, $j = 1, 2, \ldots, p$, are the centres of the basis functions;

$\boldsymbol{\sigma}_j = (\sigma_{j1}, \sigma_{j2}, \ldots, \sigma_{jd})^T$, $j = 1, 2, \ldots, p$, are the widths, which are called scaling factors for the radii $\|\boldsymbol{v} - \boldsymbol{c}_j\|$, $j = 1, 2, \ldots, p$ of the basis functions, respectively; and $\phi(\cdot)$ is a nonlinear function that monotically decreases (or increases) as $\boldsymbol{v}$ moves away from $\boldsymbol{c}_j$. In order to simplify the notation, coordinate axes-aligned Gaussian RBF functions are used. When a $1 - D$ Gaussian RBF is centred at the centroids $\boldsymbol{c}_j$, it follows from equation (B.15) that

$$\hat{g}(\boldsymbol{v}, \boldsymbol{x}) = \sum_{j=1}^{p} w_j \exp\left(-\frac{\|\boldsymbol{v} - \boldsymbol{c}_j\|^2}{2\boldsymbol{\sigma}_j^2}\right)$$

where $\boldsymbol{x}$ can be written as:

$$\boldsymbol{x}^T = \left(\boldsymbol{w}^T, \boldsymbol{\sigma}_1^T, \boldsymbol{c}_1^T, \ldots, \boldsymbol{\sigma}_j^T, \boldsymbol{c}_j^T, \ldots, \boldsymbol{\sigma}_p^T, \boldsymbol{c}_p^T\right)^T$$
$$= (w_1, \ldots, w_p, \sigma_1, c_1, \ldots, \sigma_j, c_j, \ldots, \sigma_p, c_p)^T. \tag{B.16}$$

The network can be trained to approximate $g(\boldsymbol{v})$ by finding the optimal vector $\boldsymbol{x}$ given a (possibly noisy) training set:

$$V = \left\{(\boldsymbol{v}_n, y_n) \mid n = \{1, 2, \ldots, N\}, \boldsymbol{v}_n \in \mathbb{R}^d, y_n \in \mathbb{R}\right\}.$$

Training is usually decomposed into two phases. First a clustering algorithm is used to estimate the Gaussian RBF parameters (centres and variances). Afterwards, linear supervised training can be used to estimate the weights from the hidden RBF to the output.

B.2.4   Learning Vector Quantization Neural Network

Assume that a finite training set $\{(\mathbf{x}^i, y_i), \mathbf{x} \in \mathbb{R}^n, y \in \{1, \ldots, C\} \, i = 1, \ldots, m\}$ is given. Classes are an enumerated set $\{1, \ldots, C\}$ and $\mathbb{R}^n$ denotes the potentially high-dimensional data space. Denote by $X = \{\mathbf{x}^i \mid i = 1, \ldots, m\}$ all input signals of the training set. Components of a vector $\mathbf{x} \in \mathbb{R}^n$ are referred to by subscripts, i.e., $\mathbf{x} = (x_1, \ldots, x_n)$. Learning vector quantization (LVQ) as introduced by Kohonen [95] represents every class c by a set $W(c)$ of weight vectors (prototypes) in $\mathbb{R}^n$. Weight vectors are denoted by $\mathbf{w}^r$ and their respective class label is referred to by c_r. A new signal $\mathbf{x} \in \mathbb{R}^n$ is classified by the winner-takes-all rule by an LVQ network, i.e.,

$$\mathbf{x} \longmapsto c(\mathbf{x}) = c_r \text{ such that } d(\mathbf{x}, \mathbf{w}^r) \text{ is minimum,} \tag{B.17}$$

where $d\left(\mathbf{x}, \mathbf{w}^r\right) = \left\|\mathbf{x} - \mathbf{w}^r\right\|^2 = \sum_{i=1}^{n}\left(\mathbf{x}_i - \mathbf{w}_i^r\right)^2$ denotes the squared Euclidean distance of the data point $\mathbf{x}$ to the prototype $\mathbf{w}^r$. The respective closest prototype $\mathbf{w}^r$ is called winner or best matching unit. The subset

$$\Omega = \left\{\mathbf{x}^i \in X \mid d\left(\mathbf{x}^i, \mathbf{w}^r\right) \text{ is minimum}\right\},$$

is called receptive field of neuron $\mathbf{w}^r$.

The training algorithm of LVQ aims at minimizing the classification error on the given training set. i.e., the difference of the points belonging to the th class, $\left\{x^i \in X \mid y_i = c\right\}$, and the receptive fields of the corresponding prototypes, $\cup_{w^r \in W(c)} \Omega_r$, is minimized by the adaptation process. Training iteratively presents randomly chosen data from the training set and adapts the respective closest prototype by Hebbian learning in the following way: if a vector x^i is presented, the update rule for the winner w^r has the form:

$$\triangle w^r = \begin{cases} \epsilon \cdot \left(x^i - w^r\right) & \text{if } c^r = c\left(x^i\right) \\ -\epsilon \cdot \left(x^i - w^r\right) & \text{otherwise} \end{cases}$$

$\epsilon \in (0, 1)$ is an appropriate learning rate. As explained in [144] this update can be interpreted as a stochastic gradient descent on the cost function:

$$\text{Cost}_{\text{LVQ}} = \sum_{x^i \in X} f_{LVQ}\left(d_{r+}, d_{r-}\right),$$

where d_{r+} denotes the squared Euclidean distance of $\mathbf{x}^i$ to the closest prototype $\mathbf{w}^{r^+}$ labeled with $c_{r+} = y_i$ and d_{r-} denotes the squared Euclidean distance to the closest prototype $\mathbf{w}^{r^-}$ labeled with a label c_{r-} different from y_i. For standard LVQ, the function is:

$$f_{\text{LVQ}}\left(d_{r+}, d_{r-}\right) = \begin{cases} d_{r+} & \text{if } d_{r+} \leq d_{r-} \\ -d_{r-} & \text{otherwise} \end{cases}$$

Obviously, this cost function is highly discontinuous and instabilities arise for overlapping data distributions.

Various alternatives have been proposed which substitute the training rule of LVQ by another one to achieve a more stable behavior of training also in case of overlapping classes or noisy data. Kohonen's LVQ2.1 optimizes the cost function which is obtained by setting in the above sum $f_{\text{LVQ2.1}}\left(d_{r+}, d_{r-}\right) = I_w\left(d_{r+} - d_{r-}\right)$ whereby I_w yields the identity within a window in which adaptation of LVQ2.1 takes place, and I_w vanishes outside. Still this choice might produce an instable

dynamic, and the window within which adaptation takes place must be chosen carefully.

Appendix C

Neuroimage experimental data bases

C.1 Introduction

The goal of this appendix is to give an impression of the current status of the neuroscience databases of structural and functional techniques available through Internet, with an emphasis in Magnetic Resonance Imaging (MRI). Modern neuroimaging techniques such as structural (Computed Tomography (CT), MRI and Diffusion Tensor Imaging (DTI)) or functional techniques (Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), functional Magnetic Resonance Imaging (fMRI) and functional Diffusion Tensor Imaging (fDTI)) play an important role in the diagnosis of neurodegenerative diseases.

In recent years MRI has become one of the most popular techniques used in radiology to visualize the structure and function of the body, because it is a non-ionizing radiation medical imaging technique. It provides detailed images of the body in any plane and techniques based on the principles of MRI like fMRI, DTI or fDTI are being increasingly used in the preclinical study of certain neurodegenerative diseases.

The availability of public image databases for experimental purposes allows the validation of propositions of computational methods under a common experimental framework. They allow also to reproduce the results claimed by the research groups, both relative to diagnostic issues and to computational methods. In this regard, the simulated MRI images from the BrainWeb site [4], and the clinical images from the Internet Brain Segmentation Repository (IBSR) [8], which are provided with expert segmentations that can be used as the ground thruth for

validation processes, have been widely used as benchmarks for a number of algorithms devoted to segmentation, filtering and correction of artifacts in MRI, such as the Intensity Inhomogeneity (IIH). A number of new resources have been added in recent years, the fruit of public funded ongoing research projects, to those early public database efforts. During last years new projects have been developed individually by research groups as the Laboratory of Neuro Image (LONI) [9] or through collaborations with other groups, which are working in the same research area related to image analysis and the study of neurodegenerative diseases, building consortiums such as [2, 7]. Resulting from these projects there are many public resources (images, clinical data, demographics and results of the studies) that are available for validation and refutation purposes of both clinical conclusions and computational algorithms, keeping pace with the fast evolution of the imaging devices and techniques. In fact, the filed is suffering such an explosive growth of public resources and an effervescence of results, techniques and publications that the present account may well be outdated in a very short time. The works of the PhD candidate have profited from some of these databases, namely the IBSR, BrainWeb and OASIS repositories.

C.2 IBSR

The Internet Brain Segmentation Repository (IBSR) [8] is a repository of magnetic resonance (MR) brain images and segmentation results. The IBSR was initially created in April 1996 and is maintained by Andrew Worth at the CMA. Currently there are six MR brain data sets, which were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at [8]. Most of them have T1-weighted MR images of healthy subjects. Two data sets have images of two diferent patients with brain tumors.

There are three different directories in which data sets can be found organized into "img", "seg" and "otl" directories, which contain the raw, segmented and outlined images, respectively. However not all data sets have the "seg" and/or "otl" directories. Raw images are 256x256 of 16-bit, but some of them have also these images scaled to 8-bit. Segmented images are "trinary" images (pixels are labeled as a gray matter (GM) or white matter (WM) tissue or as other), with the same dimensionality as the raw images. Outlined images are the result of semi-automated segmentation techniques performed by an expert and contain lists of points that define certain structures in each scan image. They are defined in a 512x512 grid, because they were created using oversampled images to double size. The differ-

ence with the segmentation directory is that in the trinary files group all structures are labeled into GM/WM/other, while the otl files list each neuroanatomical structure separately, so that the information provided by the segmentation trinary images is only a small subset of the information in outlined images. The data was intended to test MRI supervised and unsupervised segmentation algorithms.

Normal subject, 'Ideal' registered multi-echo brain scan

Data set 657 was created in 1996. It contains seven different image types of the same normal subject (conventional T1, PD and T2 Spin echo sequences; Fast T1, low signal/noise, SPGR (1 avg); Fast T1, better signal/noise, SPGR (2 avgs); Fast PD and T2 FSE sequence). Each of the volumes are registered scans of 18 slices ("img" format with no header information), 2.3 Mb (15.75 Mb total). Scans were acquired at the NMR Center of the Massachusetts General Hospital with a 1.5 tesla General Electric Signa. No segmentation information is provided.

Adult Male

Data set 788_6 was created in 1996. It contains T1-weighted MRI data with complete expert segmentations (trinary and outlines) from a 55 year old male subject. Each volume has been stored in 60 files that represent the slices through the brain, without header information. Scans were acquired with a 1.5 Tesla General Electric Signa. Contiguous 3.0 mm three-dimensional coronal T1-weighted spoiled gradient echo (SPGR) images of the entire brain was obtained with the following parameters: TR= 40 msec, TE= 5 msec, flip angle =40 degrees, field of view =24 cm, matrix =256x256, and averages =1.

Images were positionally normalized by imposing a standard three-dimensional brain coordinate system on each 3D MR scan [158, 55]. The repositioned scans are then resliced into normalized 3.0 mm coronal, 1.0 mm axial, and 1.0 mm sagittal scans which are used for subsequent analyses.

GM/WM segmentation (other=0; GM=128 and WM=254) was performed with a semi-automated intensity contour mapping algorithm [90] and also using signal intensity histograms. Neuroanatomical regions of interest for GM/WM segmentation include cortical GM, subcortical WM, lateral, third and fourth ventricles, caudate, putamen, globus pallidus, hippocampus-amygdala complex, thalamus proper (including all thalamic nuclei except the lateral and medial geniculate bodies), ventral diencephalic complex (including hypo-, epi-, and subthalamus, substantia nigra, red nucleus medial and lateral geniculate bodies), brainstem, cere-

bellum cortex and cerebellar central mass, according to the anatomic definitions of [56], with one exception. The central gray nuclei was subdivided at the hypothalamic fissure into thalamus proper and ventral diencephalon.

5 year old child:

Data set 1320_2 was created in 1996. It contains T1-weighted MRI data with complete expert segmentations (trinary and outlines) from a 5 year old subject. Each volume has been stored in 128 files that represent the slices through the brain, without header information. The MRI scan was acquired with a 1.5 Tesla General Electric Signa. Contiguous 1.5 mm three-dimensional coronal T1-weighted spoiled gradient echo (SPGR) images of the entire brain was attained with the following parameters: TR = 40 msec, TE = 5 msec, flip angle = 40 degrees, field of view = 24cm, matrix = 256x256, and averages = 1.

Images were positionally normalized by imposing a standard three-dimensional brain coordinate system on each 3D MR scan [158, 55]. The repositioned scans are then resliced into normalized 1.5 mm coronal, 0.9375 mm axial, and 0.9375 mm sagittal scans which are used for subsequent analyses.

GM/WM segmentation (other=0; GM=128 and WM=254) was performed following the procedure described in section C.2.

20 Normal Subjects

Data set 20_Normal was created in 1997. It contains T1-weighted MR images, from 20 normal subjects, of 3.1mm slice thickness (16-bit data; 8-bit scaled 3D data and 8-bit scaled 3D data (brain regions only)) and expert segmentations (other=0; csf=128; GM=192; WM=254). Volumes of 16-bit have been stored in 60 ".img" files that represent the slices through the brain. On the other hand 8-bit and segmented images have been stored in data files (.buchar) and header files (.hdr), where header files have four ascii numbers that give the size of the data set. The segmented and 8-bit images have less slices than the 16-bit image data. The matching segmentation must be done using the offsets given at [8] for this data set.

Images are coronal three-dimensional T1-weighted spoiled gradient echo MRI scans, that were obtained on two different imaging systems. Ten FLASH scans performed on a 1.5 tesla Siemens Magnetom MR System (Iselin, NJ) with the following parameters: TR = 40 msec, TE = 8 msec, flip angle = 50 degrees, field of view = 30 cm, slice thickness = contiguous 3.1 mm, matrix = 256x256, and

averages = 1. Ten 3D-CAPRY scans performed on a 1.5 tesla General Electric Signa MR System (Milwaukee, WI), with the following parameters: TR = 50 msec, TE = 9 msec, flip angle = 50 degrees, field of view = 24 cm, slice thickness = contiguous 3.0mm, matrix = 256x256, and averages = 1.

Images were positionally normalized by imposing a standard three-dimensional brain coordinate system on each 3D MR scan [158, 55]. The repositioned scans are then resliced into normalized 3.0 mm coronal, 1.0 mm axial, and 1.0 mm sagittal scans which are used for subsequent analyses.

Segmentation was performed on the positionally normalized scan by trained investigators using a semi-automated intensity contour mapping algorithm [90] and also using signal intensity histograms. Other neuroanatomical structures were segmented similarly [56].

Segmentation Performance Index

IBSR facilitates segmentation comparisons of six classification methods tested over this collection of images, provided by Jagath C. Rajapakse and SPM5 GM segmentation done by On Tsang. Results from Rajapakse are partially based on the method described in [127], where the comparison metric is the average overlap also called Tanimoto coefficient [48], eq. C.1:

$$T = \frac{|A \cap B|}{|A \cup B|},$$
(C.1)

where A and B are sets of voxels corresponding to different segmentations.

Tumor patients: various scans over time

Subject 126

This data set was created in 1999. It contains multiple scans of a 59 year old female with a tumor, taken at roughly 6 month intervals over three and a half years. The T1 + Gadolinium MRI scans were acquired with a 1.5 Tesla General Electric Signa and different parameters.

Subject 536

This data set was created in 1999. It contains multiple scans of a patient with a tumor (images and outlines), taken at roughly 6 month intervals over three and a

half years. Each series has been stored in 60 .img files with no header information. The pixel resolutions on these are 0.9375 x 0.9375 mm in-plane by 3.1 mm slice thickness. The outline files include 4 outlines: a contralateral reference region (the cerebral hemisphere of the right (unaffected) hemisphere), and three outlines of the enhancing tumor based upon intensity countours 1, 2 and 3 standard deviations above the mean of the contralateral reference region.

Images were registered using the CMA's standard positional normalization coordinate system.

IBSR V2.0

This data set was created in 2003 and 2004 and currently contains T1-weighted MR Image data from eighteen subjects, with expert segmentations of 43 individual structures (1.5mm slice thickness). Data are in CMA and analyze formats. For each subject there is T1-weighted volumetric images that have been 'positionally normalized' into the Talairach orientation (rotation only) and also have been processed by the CMA 'autoseg' biasfield correction routines.

GM/WM segmentation include segmentation of the 3rd ventricle, 4th ventricle, brain stem, and left and right: accumbens area , amygdala, amygdala anterior, caudate, cerebellum cortex, cerebellum exterior, cerebellum white matter, cerebral cortex, cerebral exterior, cerebral white matter, hippocampus, inf. lat. vent., lateral ventricle, pallidum, putamen, thalamus proper, ventralDC, and vessel.

C.3   BrainWeb: Simulated Brain Database

This simulated brain database (SBD) [4] was provided by McConnell Brain Imaging Centre at the Montréal Neurological Institute [17], McGill University. It contains a set of realistic MRI data volumes produced by an MRI simulator [97, 98, 40]. Currently contains simulated brain MRI data based on two types anatomical models [41] ("phantoms"): normal and multiple sclerosis (MS), which can serve as the ground truth for any analysis procedure. These anatomical models consist of a set of 3-dimensional "fuzzy" tissue membership volumes, one for each tissue class. The voxel values in these volumes reflects the proportion of tissue present in that voxel, in the range $[0, 1]$. The volumes are defined at a 1mm isotropic voxel grid in Talairach space, with dimensions 181x217x181 (XxYxZ) and origin coordinates -90,-126,-72 (x,y,z) in Talairach space.

In addition to the fuzzy tissue membership volumes, a discrete anatomical model is provided which consists of a class label (integer) at each voxel, representing the tissue which contributes the most to that voxel.

Volumes can be downloaded in MINC or raw format.

Normal Brain Volumes

Pre-computed simulated SBD

In the pre-computed SBD data are available for viewing in three orthogonal views (transversal, sagittal, and coronal) and for downloading with the parameter settings fixed to 3 modalities, 5 slice thicknesses, 6 levels of noise, and 3 levels of intensity non-uniformity. These simulations are based on an anatomical model of normal brain, which is available at a resolution of 1mm^3 and also for thicker slices (in Z direction): 3mm, 5mm, 7mm and 9mm. Tissue classes available for this phantom are: Background, CSF, GM, WM, Fat, Muscle/Skin, Skin, Skull, Glial Matter and Connective.

Custom MRI simulations interface

Through the BrainWeb custom MRI simulations interface it is possible to choose arbitrary parameters and obtain different volumes based on the same normal anatomical model as in subsection C.3.

20 sets of simulated data with specific parameters

Currently, it is only possible to download 20 different sets of T1-weighted simulated data, based on 20 anatomical models of 20 normal brains, with these specific parameters: SFLAH (spoiled FLASH) sequence with TR=22ms, TE=9.2ms, flip angle=30 deg and 1mm isotropic voxel size. Tissue classes available for these phantoms are: Background, CSF, GM, WM, Fat, Muscle/Skin, Skull, Blood vessels, Connective (region around fat), Dura Matter and Bone Marrow.

MS Lesion Brain Volumes

Pre-computed simulated SBD

In the pre-computed SBD data are available for viewing in three orthogonal views (transversal, sagittal, and coronal) and for downloading with the parameter set-

tings fixed to 3 modalities, 5 slice thicknesses, 6 levels of noise, and 3 levels of intensity non-uniformity. These simulations are based on an anatomical model of a human brain with "moderate" MS lesions, which is available at a resolution of $1mm^3$. Tissue classes available for this phantom are: Background, CSF, GM, WM, Fat, Muscle/Skin, Skin, Skull, Glial Matter, Connective and MS lesion.

Custom MRI simulations interface

Through the BrainWeb custom MRI simulations interface it is possible to choose arbitrary parameters and obtain different volumes based on three different MS anatomical models with "moderate" (the same anatomical model as subsection C.3), "mild" and "severe" MS lesions.

C.4 OASIS

The Open Access Series of Imaging Studies (OASIS) [110] provides brain imaging data that are freely available for distribution and data analysis [12]. It is made available by Dr. Randy Buckner at the Howard Hughes Medical Institute (HHMI) at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical Informatics Research Network (BIRN). Currently available data set consists of a cross-sectional collection of 416 subjects covering the adult life span aged 18 to 96 including 100 individuals over the age of 60, who have been diagnosed with early-stage Alzheimer's Disease (AD).

For each subject 3-4 T1-weighted magnetization prepared rapid gradient-echo (MP-RAGE) MRI scans were acquired on a a 1.5-T Vision scanner (Siemens, Erlangen, Germany), corresponding to multiple repetitions of the same structural protocol within a single session; a motion-corrected coregistered average of all available data; a gain-field corrected atlas-registered image to an standard space [158, 29]; and a masked version of the atlas-registered image, a GM/WM/CSF segmented image [178]. Additionally, for 20 of the nondemented subjects, images from a subsequent scan session are also included as a means of assessing acquisition reliability. All images are in 16-bit big-endian Analyze 7.5 format.

This data set has been based on the following publications for demographic [137] (gender, handedness, age, education and socioeconomic status), clinical assessments [116, 137] (Mini-Mental State Examination (MMSE) and Clinical Dementia Rating (CDR)), and derived anatomic measures [29, 58] (total intracranial

volume (eTIV), atlas scaling factor (ASF) and normalized whole brain volume (nWBV)).

The database has a complex structure, shown in figure C.1. It contains, besides the raw MRI scans, the anatomic information, the registered images, the segmented and skull stripped images. Therefore, the database can be used to test several algorithms at different points in the processing pipeline. Figure C.2 contains a table with the name codification and corresponding image types. Finally, figure C.3 contains a summary of the database demographic information.

C.5 MORPHDTI_P0001

This data set contains high SNR DTI data and the co-registered DTI data for a healthy male volunteer scanned on three separate scanning sessions over 2 days [54, 99]. Fifteen DTI scans were performed in each scan session, producing 45 DTI datasets in total. Data was acquired on a 1.5T Philips MR unit at the F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Johns Hopkins University.

C.6 MIRIAD

Multisite Imaging Research In the Analysis of Depression (MIRIAD) data set [3, 6] has dual-echo MRI scans (currently 100 subjects) acquired at Duke University, which have been anonymized and uploaded by the Neuropsychiatric Imaging Research Laboratory (NIRL) [10] to the Biomedical Informatics Research Network/Storage Resource Broker (BIRN/SRB) [2] where they are accessed at BWH (Surgical Planning Laboratory, SPL) [14] and UCLA (LONI) [9].

At the Laboratory of Neuro Imaging (LONI) a study-specific atlas is constructed from the MRI scans with both PD and T2 contrasts. Subjects' scans are linearly aligned to the study space to acquire spatial normalization factors, and a BWH tissue and structure probability field atlas is nonlinearly aligned to each individual subject. Dual-echo scans are segmented utilizing the individual subject-aligned tissue probability atlases at the Surgical Planning Laboratory, SPL; regions of interest and cerebral tissues are classified by an expectation maximization algorithm.

LONI completes the scan processing with a measurement of lobar volumes via a nonlinear registration of study-specific lobar atlas to native subject spaces, and
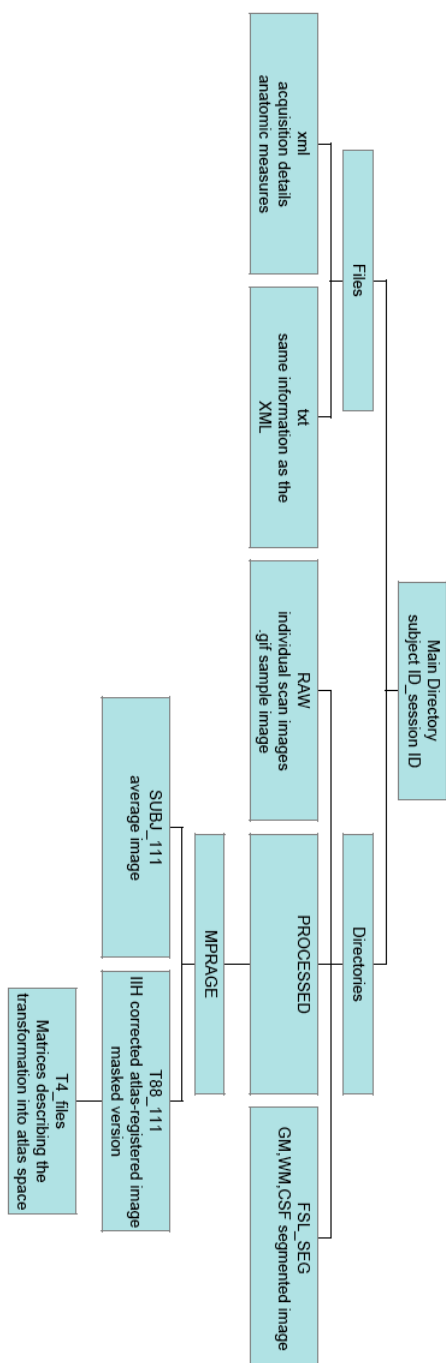
Figure C.1: File structure of the OASIS database

Name	Description	Dimensions	Vox. size	Orient
OAS1_xxxx_MRy_mpr-z_anon	Individual scan (z=repetition)	256x256x128	1x1x1.25	Sag
OAS1_xxxx_MRy_mpr_ni_anon _sbj_111	Image averaged across scans (i=# of scans)	256x256x160	1x1x1	Sag
OAS1_xxxx_MRy_mpr_ni_anon _111_t88_gfc	Gain-field corrected atlas registered average	176x208x176	1x1x1	Trans
OAS1_xxxx_MRy_mpr_ni_anon _111_t88_masked_gfc	Brain-masked version of atlas registered image	176x208x176	1x1x1	Trans
OAS1_xxxx_MRy_mpr_ni__anon 111_t88_masked_gfc_fseg	Brain tissue segmentation	176x208x176	1x1x1	Trans

Figure C.2: Types of Images included in the dataset

		Non-Demented				Demented				
Age Group	N	n	mean	male	female	n	mean	male	female	CDR 0.5/1/2
<20	19	19	18.53	10	9	0		0	0	0/0/0
20s	119	119	22.82	51	68	0		0	0	0/0/0
30s	16	16	33.38	11	5	0		0	0	0/0/0
40s	31	31	45.58	10	21	0		0	0	0/0/0
50s	33	33	54.36	11	22	0		0	0	0/0/0
60s	40	25	64.88	7	18	15	66.13	6	9	12/3/0
70	83	35	73.37	10	25	48	74.42	20	28	32/15/1
80s	62	30	84.07	8	22	32	82.88	13	19	22/9/1
≥90	13	8	91.00	1	7	5	92.00	2	3	4/1/0
Total	416	316		119	197	100		41	59	70/28/2

Figure C.3: Demographic summary of the OASIS database

regional and lobar tissue volumes are computed. The image processing results, both the images and volumetric measurements, are uploaded to the SRB. Initial statistical analyses were performed at Duke, as the associated metadata (e.g., age, sex, diagnosis, clinical scales) resided there at the beginning of the project.

C.7 ELUDE

The Efficient Longitudinal Upload of Depression in the Elderly (ELUDE) data set is an anonymized collection of a longitudinal study of late-life depression at Duke University. There are 281 depressed subjects and 154 controls included. An MR scan of each subject was obtained every 2 years for up to 8 years (total of 1093 scans). Clinical assessments occurred more frequently and consists of a battery of psychiatric tests including several depression-specific tests such as the HAM-D, CESD, and MADRS.

C.8 Alzheimer's CATX

This Data set [15] contains activation maps of 9 AD patients & 9 elderly controls [145]. It was provided by Andrew Saykin, Dartmouth Medical School and the last updated was in May 10, 2004. The task is about semantic processing: category exemplar (CATX)-identify word pairs with correct category examplar relationships from among incorrect ones.

C.9 Realistic MRI data set

This Data set [15] contains realistic brain lesion distributions generated using a lesion-deficit simulator with spatial statistical model conforming to the Frontal Lobe Injury in Childhood Study [112, 70].

C.10 DTMRI Data

This is a DTI data set [5] acquired under Human Brain Project and National Research Resource Center grant. It contains raw and processed DTI data of 15 normal population, WM atlases, DTI software. Currently the database has 2.5 mm

isotropic resolution images and 2.2 mm isotropic resolution images. Only 2.5 mm data are available.

C.11 BIRN (Biomedical Informatics Research Network)

The Biomedical Informatics Research Network (BIRN) [2] was launched in 2001 with the goal of fostering large-scale collaborations in biomedical science by utilizing emerging cyberinfrastructure. An essential feature of the project is the collaboration of computer scientists and biomedical researchers from different research disciplines to design and implement a distributed architecture of shared resources usable by all biomedical researchers in order to advance the diagnosis and treatment of disease.

FBIRN_Traveling_Subject2003

This dataset includes five healthy subjects imaged twice at each of ten FBIRN MRI scanners on successive days. Functional and structural imaging, behavioral, and demographic data are available from 100 scanning sessions on these subjects.

BrainScape_BS002

This dataset includes seventeen healthy subjects with four resting state fixation scans plus one T1 scan and one T2 scan. The data were collected as part of a study on the behavioral effects of spontaneous BOLD fluctuations [60].

BrainScape_BS003

This dataset includes ten healthy subjects scanned 3 times with 3 conditions: eyes open, eyes closed, and fixating in addition to two anatomical scans (T1 and T2) [59].

fBIRN PhaseII

The Phase II multi-site clinical imaging study consists of approximately 250 subjects, both chronic schizophrenics and age- and gender- matched controls. The

MRI data include structural and fMRI images from two separate scanning visits for each subject, including the Sternberg Item Recognition Paradigm and the Auditory Oddball paradigm, a breath hold task, and a sensorimotor task. The clinical assessments include behavioral measures, handedness and demographic measures, SES, smoking measurements, North American Adult Reading Test (NARRT), and clinical severity assessments for the clinical subjects. Currently data from three sites are released to the general research community; the remainder are awaiting IRB approvals for public data sharing.

C.12 ADNI (Alzheimer's Disease Neuroimaging Initiative)

The ADNI [1] was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership.

The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Currently there are available more than 32,000 MR and PET scans.

The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years.

C.13 Functional Brain Imaging of Young, Nondemented, and Demented Older Adults

This fMRI data set is available on Internet [16] since the year 2000. A paradigm involving repeated presentation of sensory-motor response trails was administered to 41 participants (14 young adults (18-24), 14 nondemented older adults (66-89) and 13 demented older adults (68-83)) [28].

All subjects were right-handed, english speakers, with normal (corrected) visual acuity. A history of neurological or visual illness served as exclusion criteria for all potential subjects. Furthermore, older adults were excluded if they had

neurologic, psychiatric or mental illness which could cause dementia. Dementia status was determined using recruitment and assessment procedures developed by the Alzheimer's Disease Research Center at Washington University. Clinicopathology studies in cognitively healthy aging and AD; Relation of histologic markers to dementia severity, age, sex, and APOE genotype; CDR.

Stimulus display was controlled by a Power Macintosh computer (Apple, Cupernico, CA) using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). Keypress responses were recorded using a fiber-optic light-sensitive keypress connected to a PsyScope button box (Carnegie Mellon University, Pittsburgh, PA). All buttons except one were physically covered to minimize response complexity.

Stimuli were rear projected (Am- Pro Model LCD-150, Ampro, Melbourne, FL) onto a screen placed at the back of the magnet bore. Participants viewed the screen through a mirror fastened to the top of the head coil. Participants requiring corrective lenses (mostly older adults) were supplied magnetcompatible glasses.

The basic task paradigm consisted of presentation of a 1.5-sec duration visual stimulus. Participants pressed a key with their right index fingers upon stimulus onset. The visual stimulus was an 8-Hz counterphase flickering (black to white) checkerboard subtending approximately 12° of visual angle (6° in each visual field). Stimulus parameters were identical to those used by [114]. The stimulus onset was triggered at the beginning of the image acquisition via the PsyScope button box. Spatial frequency of the checkerboard decreased with visual angle to be approximately constant in relation to acuity across the visual field. Runs were structured such that for every eight-image acquisition (21.44 sec), one of two kinds of trial condition were presented (15 trials per run for a total of 60 trials per subject).

Task trials either involved stimuli presented in isolation (one-trial condition) or in pairs with an inter-trial interval of 5.36 sec (two-trial condition). One-trial and two-trial conditions were pseudorandomly intermixed such that eight trials of one type and seven of the other appeared in each run. The logic of this design [42] is that the onetrial conditions can be examined to determine the evoked hemodynamic response to an isolated, transient event. The two-trial conditions further allowed the summation properties of the hemodynamic response to be examined: To the degree that the added responses in the two-trial conditions were similar to the responses in the one-trial conditions, the hemodynamic dynamic response exhibits linear summation. Four image acquisitions involving only fixation were acquired prior to the first trial and following the last trial in each run.

C.14 Neuroscience Database Gateway (NDG)

The Neuroscience Database Gateway (NDG) [11] began in 2004 as a pilot project developed by the Society's Brain Information Group (BIG). The NDG is now overseen by the Society's Neuroinformatics Committee and is hosted at Yale University (by Gordon Shepherd and Luis Marenco). The SfN Neuroscience Database Gateway provides links to five main types of database: Databases of experimental data; knowledge bases; software tools for neuroscience; bioinformatics resources; providers of research materials; all neuroscience databases.

C.15 LONI Image Data Archive (IDA)

The LONI Image Data Archive (IDA) [9] is a user-friendly environment for archiving, searching, sharing, tracking and disseminating neuroimaging and related clinical data. The IDA is utilized for dozens of neuroimaging research projects across North America and Europe and accommodates MRI, PET, MRA, DTI and other imaging modalities. A flexible data de-identification engine and encrypted file transmission help ensure compliance with patient-privacy regulations.

C.16 mBIRN Data Repository (mBDR)

The mBIRN Data Repository (mBDR) [3] is a public resource presented by the Morphometry testbed of the Biomedical Informatics Research Network (BIRN). It includes a range of raw and post-processed MRI images, related derived measures, and related subject measures. The data are organized by project OASIS, morphDTI_p0001, MIRIAD, ELUDE

C.17 fMRI Data Center (fMRIDC)

The fMRIDC is a public repository [16] of peer-reviewed fMRI publications projects and their underlying data. Currently there are 122 data sets available.

C.18   DEnLab Data Repository

The repository [15] contains both medical data such as the ones referred in se-
cions C.8 and C.9, as well as some information about methods and computational
techniques for medical image processing.

C.19   ICBM Human Atlases

The LONI Atlas site [9] consists of a collection of data, online viewers, images
and animations that describe the various atlases (e.g. Alzheimer's disease Tem-
plate) developed at LONI. A complete description and discussion of these atlas
requires a dedicated chapter.

Appendix D

The General Linear Model and SPM

In this appendix we give a brief account of the fundamentals of the statistical methods that are implemented in SPM software, freely available from the website `http://www.fil.ion.ucl.ac.uk/spm/`, and constitute the standard analysis protocol for several applications, among them the Voxel Based Morphometry (VBM) and the analysis of fMRI data. We introduce the nomenclature, we review the General Linear Model (GLM) formulation, the statistical inference tests and the way Random Field Theory are applied to obtain detection of significative differences (VBM) or activations (fMRI).

D.1 Introduction

Characterizing a regionally specific effect rests on estimation and inference. Inferences in neuroimaging may be about differences expressed when comparing one group of subjects to another or, within subjects, changes over a sequence of observations. They may refer to structural differences (e.g. in VBM) [20] or neurophysiological measures of brain functions (e.g. fMRI or functional magnetic resonance imaging). The principles of data analysis are very similar for all of these applications.

Looking for any effect in each voxel we fit a statistical model to the data, to give us model parameters. We then use the model parameters to look for an effect we are interested in, such as the difference between a task and baseline or identifying and characterizing structural differences among populations. To do this, we usually calculate a statistic for each brain voxel that tests for the effect of interest in that voxel.

The result is a large volume of statistic values. We now need to decide if this volume shows any evidence of the effect. To do this, we have to take into account that there are many thousands of voxels and therefore many thousands of statistic values. This is the multiple comparison problem in functional imaging. Random field theory is a recent branch of mathematics that can be used to solve this problem.

D.2 Nomenclature

- Contrast: The linear combination of the estimated parameters to formulate a question on the data as a statistical test, either a t-test or an F-test.

- Degrees of Freedom: of the distribution of the statistics specified by the contrast.

- Design Matrix: The matrix of covariates for the GLM. In the case of the fMRI analysis it corresponds to the design of the experiment and the models of diverse effects in the data. In the VBM it corresponds to the specification of the two population t-test and some covariates.

- The Euler characteristic (EC): is a property of an image after it has been thresholded (number of blobs of Z-scores above a certain threshold).

- Family Wise Error (FWE): the accepted rate of false positives when we have a large number of simultaneous tests (the probability of having even one false discovery over the ensemble of tests.)

- False Discovery Rate (FDR): the expected proportion of incorrectly rejected null hypotheses.

- Resel: is a term introduced by Worsley [172] and allows us to express the search volume in terms of the number of 'resolution elements' in the statistical map. We can use the FWHM to calculate the number of resels in the image. In 3D a resel is a cube of voxels of size (FWHM in x) by (FWHM in y) by (FWHM in z).

- Statistical parametric maps (SPMs) are images or fields with values that are, under the null hypothesis, distributed according to a known probability density function, usually the Student's t or F-distributions. These are known colloquially as t- or F-maps.

SPM Statistical parametric map(ping)	**fMRI** Functional Magnetic resonance imaging
GLM general linear model	**PET** Positron emission tomography
RFT Random field theory	**HRF** Haemodynamic response function
VBM Voxel-based morphometry	**IID** Independent and identically distributed
FWE Family-wise error	**DCT** Discrete cosine transform
FDR False discovery rate	**MR** Magnetic Resonance
MRI Magnetic resonance imaging	**FWHM** Full width at half maximum

Table D.1: Some common acronyms

- Statistical parametric mapping (SPM software) is a suit of MATLAB functions and subroutines with some externally compiled C routines. The SPM software package has been designed for the analysis of brain imaging data sequences.

D.3 The GLM Formulation

The GLM is a statistical linear model, which supports many parametric techniques such as linear or multiple regression models by an equation that expresses the observed response variable in terms of a linear combination of explanatory variables X, the *design matrix*, plus a well behaved error term. Each column of the design matrix corresponds to an effect one has built into the experiment or that may confound the results. The general expression of the GLM model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{D.1}$$

where $\mathbf{y}$ is a $M \times 1$ vector of observations and M is the number of observations. The $M \times L$ matrix $\mathbf{X}$ is the *design matrix* and it has one row per observation and one colum per model parameter (explanatory variable). $\boldsymbol{\beta}$ is a $L \times 1$ vector of unknown parameters and ε is a $M \times 1$ vector of error terms, usually assumed to be $\varepsilon_j \overset{iid}{\sim} N(0, \sigma^2)$. In some applications (i.e. fMRI) the columns of $\mathbf{X}$ are called the regressors and the β parameters are the regression coefficients. If we consider the simultaneous modeling of all the voxels in a collection of MRI data, either a a group of subject's MRI for VBM or a sequence of T2* MRI volumes for fMRI, then we have:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \boldsymbol{\varepsilon}, \tag{D.2}$$

where $\mathbf{Y}$ is the $M \times N$ matrix of voxel observations reshaped in a columnwise disposition, and M is as before the number of observations, and N is the number of voxels. The $L \times N$ matrix B is the matrix of parameters of the models of all the voxels also disposed columnwise.

But there are some situations that this model does not cover. For instance, in fMRI (a) BOLD responses have a delayed and dispersed form, (b) the BOLD signal includes substantial amounts of low-frequency noise and (c) the data are serially correlated (temporally autocorrelated) violating the statistical independence assumption about the noise model in the eq. D.1.

To account for the shape of BOLD response a convolution model is introduced, which convolves stimulus function with a canonical haemodynamic response function (HRF). The response of a linear time-invariant (LTI) system is the convolution of the input with the system response to an impulse.

The solution for the low-frequency noise is the highpass filtering. Removing low frequencies from the time-series allows the model to fit serial correlations over a more restricted frequency range or shorter time spans. Drift removal can be implemented by including drift terms in the design matrix or by including the implicit residual forming matrix of DCT set in S to make it a highpass filter, $S\boldsymbol{y} = SX\boldsymbol{\beta} + \varepsilon$.

To deal with serial correlations Worsley and Friston [173] proposed a procedure known as $pre - colouring$, that replaces the unknown endogenous autocorrelation by imposing some known autocorrelation structure on the data (filtering with a matrix W) and use Satterthwaite correction for df's (degrees of freedom). An alternative approach [173] is the $pre - whitening$, which tries to estimate a filter matrix W from the data to de-correlate the errors. $W\boldsymbol{y} = WX\boldsymbol{\beta} + W\boldsymbol{\varepsilon}$.

D.4 Parameter estimation

In this step we need some method of estimating parameters that "best" fits the data. In the case of SPM5 that is achieved by the method of *ordinary least squares*, which defines "best" as the sum, S, of squared residuals (or differences between the actual and fitted values):

$$S = \sum_{m=1}^{M} e_m^2 = \sum_{m=1}^{M} (y_m - x_{m1}\hat{\beta}_1 - ... - x_{mL}\hat{\beta}_L)^2 \tag{D.3}$$

where $e = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}} = \left[\hat{\beta}_1,..., \hat{\beta}_L\right]$ is the vector of parameter

estimates. The minimum of S is found by equating the gradient to zero:

$$\frac{\partial S}{\partial \hat{\beta}_l} = 2\sum_{m=1}^{M}(-x_{ml}\hat{\beta}_1 - ... - x_{mL}\hat{\beta}_L) = 0, \; l = 1, ..., L \qquad \text{(D.4)}$$

The least squares estimate $\hat{\boldsymbol{\beta}}$ satisfy the normal equations $\mathbf{X}^T\mathbf{y} = (\mathbf{X}^T\mathbf{X})\hat{\boldsymbol{\beta}}$, thus, if $(\mathbf{X}^T\mathbf{X})$ is invertible the least squares estimates are:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \qquad \text{(D.5)}$$

If we take into account all the voxels, the parameter estimation can be done for all the voxels simultaneously:

$$\hat{\boldsymbol{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \qquad \text{(D.6)}$$

If X has linearly dependent columns, it is *rank deficient*, i.e. (X^TX) is singular and has no inverse. In this case, the model is overparameterized (overdetermined), therefore the parameters are not 'unique', 'identifiable' or 'estimable'. For such models, X^TX is not invertible so we must resort to generalised inverses (SPM uses the pseudo-inverse).

The term X^TX is called information matrix and reflects the orthogonality of the design matrix. High covariance between the columns of the design matrix introduces redundancy. This can increase the covariance of the parameter estimates $(X^TX)^{-1}$ and leads to low efficiency (depending on the particular contrast). In the SPM package, for each pair of columns of the design matrix, the orthogonality matrix depicts the magnitude of the cosine of the angle between them, with the range 0 to 1 mapped from white to black. The cosine of the angle between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is obtained by: $\cos\alpha = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}$. If both vectors have zero mean then the cosine of the angle between the vectors is the same as the correlation between the two variates.

D.5 Statistical Inference

This is the last part of the analysis where we derive t- and F-statistics to test for a linear combination of effects. The way to do that is using voxel-wise statistical tests, which are comprised in a SPM. After that, RFT is applied to correct for multiple dependent comparisons.

Using statistical hypothesis testing we look for a linear combination of effects. The null hypothesis H_0 is the hypothesis that there is no effect, and the alternative

hypothesis H_1 is the hypothesis that there is an effect. Absence of evidence is not evidence of absence, therefore one can not accept the null hypothesis (one can just fail to reject it).

The evidence about H_0 is summarised by the test statistic T, which is typically small in magnitude when the hypothesis H_0 is true and large when false. We need to know the distribution of T under the null hypothesis. We denote α the acceptable false positive rate (Type I Error) and u_α the threshold, which controls $\alpha = p(T > u_\alpha \,|\, H_0)$. We obtain the conclusion about the hypothesis through the observed significance level (or p value), which is the smallest fixed level t at which the null hypothesis can be rejected in favour of the alternative hypothesis if $t > u_\alpha$, where $p(T > t \,|\, H_0)$.

We are usually not interested in the whole $\boldsymbol{\beta}$, the vector of effects. Therefore, we define $\boldsymbol{c}$ as a contrast vector of the same length as $\boldsymbol{\beta}$ to select specific effects of interest, so that $\boldsymbol{c}^T \boldsymbol{\beta}$ is a linear combination of regression coefficients $\boldsymbol{\beta}$. Under the i.i.d. assumptions:

$$\boldsymbol{c}^T \hat{\boldsymbol{\beta}} \sim N(\boldsymbol{c}^T \boldsymbol{\beta}, \sigma^2 \boldsymbol{c}^T (X^T X)^{-1} \boldsymbol{c}). \tag{D.7}$$

In the SPM software we can perform t or F statistics for making inferences, depending on what we are looking for. While t-tests are used to assess individual effects, F-tests allow us to assess the significance of a set of effects. The t-tests are specified by one dimensional contrasts (t-contrast) so that the alternative hypothesis corresponds to the contrast being greather than 0, that is, $H_1 : \boldsymbol{c}^T \boldsymbol{\beta} > 0$. The null hypothesis is $H_0 : \boldsymbol{c}^T \boldsymbol{\beta} = 0$ and the test statistic is computed as

$$T = \frac{\boldsymbol{c}^T \hat{\boldsymbol{\beta}}}{\sqrt{var(\boldsymbol{c}^T \hat{\boldsymbol{\beta}})}} = \frac{\boldsymbol{c}^T \hat{\boldsymbol{\beta}}}{\sqrt{\sigma^2 \boldsymbol{c}^T (X^T X)^{-1} \boldsymbol{c}}} \sim t_{N-p} \tag{D.8}$$

Where T does not depend neither on the scaling of regressors nor on the scaling of the contrast.

F-test uses the extra-sum-of-squares principle. F-contrast allows to assess general linear hypotheses, and compare models in a hierarchy, where inference is based on an F-statistic. Therefore employing the extra-sum-of-squares principle we can compare the model and a reduced model. Suppose we have a model with parameter vector $\boldsymbol{\beta}$ that can be bi-partitioned into, $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T \right]^T$, and suppose we wish to test $H : \boldsymbol{\beta}_1 = 0$. The correspondig partitioning of the design matrix X is $X = \left[X_1 \vdots X_2 \right]$, and the full model is:

$$Y = \left[X_1 \vdots X_2 \right] \begin{bmatrix} \boldsymbol{\beta}_1 \\ \cdots \\ \boldsymbol{\beta}_2 \end{bmatrix} + \varepsilon, \tag{D.9}$$

which, when H is true, reduces to the reduced model $Y = X_2 \boldsymbol{\beta}_2 + \epsilon$. The residual sum-of-squares for the full model is defined as $RSS = \sum \hat{\varepsilon}^2_{full}$ and that of the reduced model as $RSS_0 = \sum \hat{\varepsilon}^2_{reduced}$. Therefore the following F-statistic expresses the ratio of explained variability and unexplained variability (error)

$$F \propto \frac{RSS_0 - RSS}{RSS} \sim F_{v_1, v_2} \tag{D.10}$$

where the extra sum-of-squares due to β_1 after β_2 is defined as $ESS = RSS_0 - RSS$; $v_1 = rank(X) - rank(X_0)$ and $v_2 = M - rank(X)$.

Some remarks about F-test:

- F tests a weighted sum of squares of one or several combinations of the regression coefficients $\boldsymbol{\beta}$.

- In practice, we do not have to explicitly separate X into $[X_1 X_2]$ thanks to multidimensional contrasts.

- Null Hypothesis: $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{n_1} = 0$; $H_1 :$ at least one $\beta_k \neq 0$.

- In testing uni-dimensional contrast with an F-test, for example $\beta_1 - \beta_2$, the result will be the same as testing $\beta_2 - \beta_1$. It will be exactly the square of the t-test, testing for both positive and negative effects.

The t-contrast basically tells us about whether a specific linear contrast of beta estimates differs from zero. The F-contrast tells us about how much a given linear contrast of parameter estimates (as a subset of such contrasts) contributes uniquely to explaining variance in the data.

D.6   Thresholding methods

In this section we will give a brief summary of the Bonferroni correction and the Random Field Theory (RFT), which correct for multiple dependent comparisons and are used by SPM2 and later versions of the SPM package. Although there are other thresholding methods that are currently implemented in many fMRI studies such as cluster thresholds, thresholds obtained by permutations and ad hoc methods.

D.6.1 Bonferroni correction

If we have a number n of simultaneous statistical tests, such as it is the case when performing VBM or fMRI analysis, for a given significance level α, the probability that all the computed statistics fall, by chance, below the corresponding threshold is $(1 - \alpha)^n$, [61]. The Family-Wise Error (FWE) rate P^{FWE} is defined as the probability that at least one of the statistics is, by chance, above the test threshold is $P^{FWE} = 1 - (1 - \alpha)^n$. Since α is small the family-wise error rate can be approximated by $P^{FWE} \leq n \cdot \alpha$. Therefore the following is a lower bound on the value of the significance level if we want to ensure that the probability of false positives is the FWE rate:

$$\alpha = \frac{P^{FWE}}{n}. \tag{D.11}$$

This is the so-called Bonferroni correction, which is very conservative when the statistical test are not in fact independent, such as in the case of the VBM and fMRI.

D.6.2 Random Field Theory

In the case of MRI image analysis the assumption of independence of the voxel-wise test is not correct, because there are always spatial correlations due to the physiological processes. The extreme conservativeness of the Bonferroni method, coupled with its inability to take into consideration the particular features of fMRI data, requires other technique for error control such as the RFT in which the expected Euler characteristic (EC) is given as an approximation of the probability of a FWE $P^{FWE} \approx E[EC]$.

In RFT we need a smooth statistical map, so its application proceeds in stages. First, we estimate the smoothness (spatial correlation) of our statistical map. Then we use the smoothness values in the appropriate RFT equation. This allows us to calculate the threshold at which we would expect 5 per cent of equivalent statistical maps arising under the null hypothesis to contain at least one area above threshold. For an image of two dimensions

$$E[EC] = R(4 \log_e 2)(2\pi)^{-\frac{3}{2}} Z_t e^{-\frac{1}{2} Z_t^2} \tag{D.12}$$

Where R is the number of resels and Z_t is the Z-score threshold. There are other kind of random fields besides of Z-scores, such as t, F and χ^2 random fields (i.e. the SPM software uses t and F random fields).

D.7 SPM for Voxel-Based Morphometry

Voxel-based morphometry (VBM) has been developed to characterise brain differences in vivo using structural magnetic resonance (MR) images. VBM is a simple procedure that enables classical inferences about the regionally-specific effects, of experimental factors, on some structural measure. These effects are tested after discounting the large-scale anatomical differences removed by spatial normalization. Because these differences have been removed, VBM is not a surrogate for classical volume analysis of large structures or lesions [113].

D.7.1 Preprocessing for VBM

The procedure is relatively straightforward, and involves spatially normalizing and segmenting high-resolution magnetic resonance images into the same stereotaxic space. The MR images are segmented, producing images that reflect the spatial distribution of a tissue type or attribute (e.g., GM). To compare brains of different subjects, all the GM segments are warped to the same stereotaxic space. A correction can be applied to the data that accounts for expansion and contraction during this nonlinear spatial normalization. The normalized segments are then smoothed. This makes each voxel a measure of the proportion of the brain in a region around the voxel that is GM (i.e. GM density).

D.7.2 Design matrix construction, experimental design

Statistical analysis using the GLM is performed on the pre-procesed images to identify regions that are significantly related to the effects under study [63]. Each row of the design matrix corresponds to a scan, whereas each column is some effect that is modelled.

Occasionally, images from different scanners or data from different sequences are mixed. The effects of different types of images can generally be modelled by including confounds in the design matrix.

D.7.3 Specific statistical assumptions/problems

Parametric statistical tests are based on assumptions about the data. In particular, they assume that the residuals after fitting the model are normally distributed. There is a special caveat in VBM concerning the normality assumption: GM segments contain values between zero and one, where most of the values are close to

either of the extremes. In this case, smoothing is essential to render the residuals sufficiently normal.

For group comparisons, normality can be assured with smoothing at or above 4mm FWHM [20]. For single-case comparisons, numerical studies suggest that at least 12mm should be used [141].

D.8 SPM for fMRI analysis

In fMRI analysis we can address two different questions of interest. With block designs, which are especially useful for detection of the "activated" voxels in response to a given task, compared to a control condition and with event-related designs, which provide a means of estimating the hemodynamic response function.

D.8.1 Preprocessing for fMRI

It is always necessary to perform some preprocessing, or cleaning, of the data, prior to statistical analysis. Different software packages accomplish this in different ways; even different laboratories using the same software preprocess their data differently [118]. However, there are certain commonalities. In SPM the general steps for the preprocessing are: motion correction (fMRI-fMRI: a) head motion and b) slice timing differences), spatial normalisation (fMRI-MRI: a) corregistration between structural and functional data for further visualization of the activations and b) align the fMRI scans to a common coordinate space) and smoothing.

D.8.2 Design matrix construction, experimental design

There are two main approaches to the design of fMRI experiments (block designs and event-related designs), from the perspective of stimulus presentation, which are capable of effectively addressing two different questions of interest.

In block designs periods of rest alternated with periods of task (t statistics), or periods of different task were alternated (F statistics). These designs are especially useful for locating voxels in which the level of activity is significantly different in the task versus the control conditions.

Event-related designs generally consist of rapidly presented interleaved trials of multiple event types of short duration, and provide a mean of estimating the hemodynamic response function.

D.8.3 Specific statistical assumptions/problems

There are many sources of noise that are prevalent in fMRI data. Even after pre-procesing, considerable variability remains (statistical noise) in data, presenting a challenge to statistical methods and limiting, to some extent, the types of analysis that will be effective. Another complication is the spatial and temporal structure in the data. fMRI data acquired on a single subject are abundant, noisy, and they are highly correlated both spatially and temporally.

Bibliography

[1] Alzheimer's Disease Neuroimaging Initiative (ADNI) at http://www.loni.ucla.edu/adni/.

[2] Biomedical Informatics Research Network: http://www.nbirn.net/.

[3] BIRN Data Repository (BDR) available at http://www.nbirn.net/bdr/index.shtm.

[4] BrainWeb available at http://www.bic.mni.mcgill.ca/brainweb/.

[5] DTI Data available at http://lbam.med.jhmi.edu/.

[6] Human Imaging Database available at http://nirlarc.duhs.duke.edu.

[7] International Consortium for Brain Mapping (ICBM): http://www.loni.ucla.edu/ICBM/.

[8] Internet Brain Segmentation Repository (IBSR) available at http://www.cma.mgh.harvard.edu/ibsr/.

[9] Laboratory of Neuro Imaging (LONI), UCLA at http://www.loni.ucla.edu/.

[10] Neuropsychiatric Imaging Research Laboratory, Departments of Psychiatry & Radiology at Duke University Medical Center: http://nirlweb.mc.duke.edu/.

[11] Neuroscience Database Gateway (NDG) at http://ndg.sfn.org/.

[12] Open Access Series of Imaging Studies (OASIS) available at http://www.oasis-brains.org/.

[13] Statistical Parametric Mapping available at http://www.fil.ion.ucl.ac.uk/spm/.

[14] Surgical Planning Laboratory at Brigham & Women's Hospital: http://www.spl.harvard.edu/.

[15] The Data Engineering Laboratory (DEnLab), Department of Computer and Information Sciences and the Center for Information Science and Technology at Temple University: http://denlab.temple.edu.

[16] The fMRI Data Center (fMRIDC) at http://www.fmridc.org.

[17] The McConnell Brain Imaging Centre (BIC) at the Montreal Neurological Institute: http://www2.bic.mni.mcgill.ca/.

[18] M.N. Ahmed, S.M. Yamany, N. Mohamed, A.A. Farag, and T. Moriarty. A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *Medical Imaging, IEEE Transactions on*, 21(3):193–199, 2002.

[19] I. Álvarez, M. López, J.M. Górriz, J. Ramírez, D. Salas-González, F. Segovia, and C.G. Puntonet. Automatic classification system for the diagnosis of Alzheimer disease using component-based SVM aggregations. *International Conference on Neural Information Processing of the Asia-Pacific Neural Network Assembly (ICONIP)*, 2008.

[20] J. Ashburner and K. Friston. Voxel-Based Morphometry–The methods. *Neuroimage*, 11(6):805–821, 2000.

[21] M.S. Atkins, K. Siu, B. Law, J.J. Orchard, and W.L. Rosenbaum. Difficulties of T1 brain MRI segmentation techniques. *Image Processing, Proceedings of SPIE*, 4684(1):1837–1844, 2002.

[22] A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.*, 7(6):1129–1159, 1995.

[23] J.C. Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981.

[24] J.C. Bezdek and S.K. Pal. Fuzzy models for pattern recognition. *IEEE Press, Piscataway, NJ*, 1991.

[25] S.M. Bhandarkar and P. Nammalwar. Segmentation of multispectral MR images using a hierarchical Self-Organizing Map. In *Proceedings of the Fourteenth IEEE Symposium on Computer-Based Medical Systems (CBMS)*, page 294, Washington, DC, USA, 2001. IEEE Computer Society.

[26] C. Brechbühler, G. Gerig, and G. Szekely. Compensation of spatial inhomogeneity in MRI based on a parametric bias estimate. In *Proceedings of the 4th International Conference on Visualization in Biomedical Computing (VBC)*, pages 141–146, London, UK, 1996. Springer-Verlag.

[27] D.S. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1998.

[28] R L Buckner, A Z Snyder, A L Sanders, M E Raichle, and J C Morris. Functional brain imaging of young, nondemented, and demented older adults. *J. Cogn. Neurosci.*, 12 Suppl 2:24–34, 2000.

[29] R.L. Buckner, D. Head, J. Parker, A.F. Fotenos, D. Marcus, J.C. Morris, and A.Z. Snyder. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, 23(2):724–738, 2004.

[30] A.L. Buffington, C.A. Hanlon, and M.J. McKeown. Acute and persistent pain modulation of attention-related anterior cingulate fMRI activations. *Pain*, 113(1):172–184, 2005.

[31] C. Burges. A tutorial on Support Vector Machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[32] P. Burrascano. Learning Vector Quantization for the Probabilistic Neural Network. *IEEE Trans. Neural Networks*, 2:458–461, 1991.

[33] V.D. Calhoun and T. Adali. Unmixing fMRI with Independent Component Analysis. *Engineering in Medicine and Biology Magazine, IEEE*, 25(2):79–90, March-April 2006.

[34] V.D. Calhoun, K.A. Kiehl, P.F. Liddle, and G.D. Pearlson. Aberrant localization of synchronous hemodynamic activity in auditory cortex reliably characterizes schizophrenia. *Biol. Psychiatry*, 55(8):842–9, Apr 2004.

[35] V.D. Calhoun, J.J. Pekar, and G.D. Pearlson. Alcohol intoxication effects on simulated driving: exploring alcohol-dose effects on brain activation using functional MRI. *Neuropsychopharmacology*, 29(11):2097–2017, Nov 2004.

[36] C.-C. Chang and C.-J. Lin. LIBSVM: a library for Support Vector Machines available at http://www.csie.ntu.edu.tw/ cjlin/libsvm/. 2001.

[37] S. Chaplot, L.M. Patnaik, and N.R. Jagannathan. Classification of magnetic resonance brain images using wavelets as input to Support Vector Machine and Neural Network. *Biomedical Signal Processing and Control*, 1(1):86–92, 2006.

[38] S. Chen, C.F.N. Cowan, and P.M. Grant. Orthogonal Least Squares Learning algorithm for Radial Basis Function Networks. *Neural Networks, IEEE Transactions on*, pages 302–309, 1991.

[39] C. Chinrungrueng and C.H. Sequin. Optimal adaptive K-means algorithm with dynamic adjustment of learning rate. *IEEE Trans Neural Netw*, 6(1):157–69, 1995.

[40] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, G.B. Pike, and A.C. Evans. BrainWeb: online interface to a 3D MRI simulated brain database. *NeuroImage*, 5:425, 1997.

[41] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging*, 17(3):463–468, 1998.

[42] A.M. Dale and R.L. Buckner. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5(5), 1997.

[43] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S.M. Resnick. Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging*, 29(4):514–523, Apr 2008.

[44] E. de Bodt, M. Cottrell, P. Letremy, and M. Verleysen. On the use of Self-Organizing Maps to accelerate vector quantization. *Neurocomputing*, 56:187–203, 2004.

[45] T. Deneux and O. Faugeras. Using nonlinear models in fMRI data analysis: model selection and activation detection. *NeuroImage*, 32(4):1669–1689, 2006.

[46] A.P. Dhawan. *Medical image analysis*. IEEE Press Series in Biomedical Engineering. John Wiley & Sons, IEEE Press Series in Biomedical Engineering, 2003.

[47] W.P. dos Santos, R.E. de Souza, A.F.D. e Silva, and P.B. Santos Filho. Evaluation of Alzheimer's disease by analysis of MR images using Multilayer Perceptrons, Polynomial Nets and Kohonen LVQ classifiers. *Computer Vision/Computer Graphics Collaboration Techniques*, pages 12–22, 2007.

[48] R.O. Duda and P.E. Hart. *Pattern classification and scene analysis*. A Wiley-Interscience Publication, New York: Wiley, 1973, 1973.

[49] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification*. Wiley Interscience, New York, second ed. edition, 2001.

[50] J.C. Dunn. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybern*, 3:32–57, 1973.

[51] P.A. Estévez and C.J. Figueroa. Online data visualization using the Neural Gas Network. *Neural Networks*, 19(6-7):923–34, 2006.

[52] Y. Fan, , S.M. Resnick, S. Wu, and C. Davatzikos. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *NeuroImage*, 41(2):277–285, 2008.

[53] Y. Fan, D. Shen, and C. Davatzikos. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 8(Pt 1):1–8, 2005.

[54] J.A.D. Farrell, B.A. Landman, C.K. Jones, S.A. Smith, J.L. Prince, P.C.M. van Zijl, and S. Mori. Effects of SNR on the accuracy and reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *Journal of Magnetic Resonance Imaging*, 2007.

[55] P.A. Filipek, D.N. Kennedy, and V.S. Caviness. Volumetric analysis of central nervous system neoplasm based on MRI. *Pediatric Neurology*, 7:347–351, 1991.

[56] P.A. Filipek, C. Richelme, D.N. Kennedy, and V.S. Caviness. The young adult human brain: an MRI-based morphometric analysis. *Cereb Cortex*, 4(4):344–360, 1994.

[57] J.C. Fort. SOM's mathematics. *Neural Networks*, 19(6-7):812–6, 2006.

[58] A.F. Fotenos, A.Z. Snyder, L.E. Girton, J.C. Morris, and R.L. Buckner. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, 64(6):1032–1039, Mar 2005.

[59] M.D. Fox, A.Z. Snyder, J.L Vincent, M. Corbetta, D.C. Van Essen, and M.E. Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A*, 102(27):9673–9678, 2005.

[60] M.D. Fox, A.Z. Snyder, J.L. Vincent, and M.E. Raichle. Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron*, 56(1):171–184, October 2007.

[61] R.S.J. Frackowiak, K.J. Friston, R.J. Dolan, C.D. Frith, C.J. Price, J.T. Ashburner, S. Zeki, and W.D. Penny. *Human brain function*. Academic Press, 2004.

[62] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *EuroCOLT '95: Proceedings of the Second European Conference on Computational Learning Theory*, pages 23–37, London, UK, 1995. Springer-Verlag.

[63] K.J. Friston, A.P. Holmes, K.J. Worsley, J.B. Poline, C. Frith, and R.S.J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

[64] G. Fung and J. Stoeckel. SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. *Knowl. Inf. Syst.*, 11(2):243–258, 2007.

[65] M. García Sebastián, E. Fernández, M. Graña, and F.J. Torrealdea. A parametric gradient descent MRI intensity inhomogeneity correction algorithm. *Pattern Recogn. Lett.*, 28(13):1657–1666, 2007.

[66] M. García Sebastián, A.I. González, and M. Graña. Derivation of SOM-like rules for intensity inhomogeneity correction in MRI. *Computational and Ambient Intelligence, IWANN 2007*, 4507:669–676, 2007.

[67] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):609–628, 1990.

[68] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):367–383, 1992.

[69] G. Gerig, O. Kubler, R. Kikinis, and F.A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Trans. Med. Imaging*, 11(2):221–32, 1992.

[70] J.P. Gerring, K.D. Brady, A. Chen, C.B. Quinn, K.J. Bandeen Roche, M.B. Denckla, and R.N. Bryan. Neuroimaging variables related to the development of secondary attention deficit hyperactivity disorder in children who have moderate and severe closed head injury. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37:647–654, 1998.

[71] R.D. Gibbons, N.A. Lazar, D.K. Bhaumik, S.L. Sclove, H.Y. Chen, K.R. Thulborn, J.A. Sweeney, K. Hur, and D. Patterson. Estimation and classification of fMRI hemodynamic response patterns. *NeuroImage*, 22(2):804–814, 2004.

[72] J.O. Glass, W.E. Reddick, C.-S. Li, F.H. Laningham, K.J. Helton, and C.-H. Pui. Computer-aided detection of therapy-induced leukoencephalopathy in pediatric acute lymphoblastic leukemia patients treated with intravenous high-dose methotrexate. *Magn. Reson. Imaging*, 24(6):785–91, Jul 2006.

[73] A.F. Goldszal, C. Davatzikos, D.L. Pham, M.X. Yan, R.N. Bryan, and S.M. Resnick. An image-processing system for qualitative and quantitative volumetric analysis of brain images. *J. Comput. Assist. Tomogr.*, 22(5):827–37, 1998.

[74] A.I. González and M. Graña. Controversial empirical results on batch versus one pass online algorithms. *Workshop On Self-Organizing Maps*, 2005.

[75] A.I. González, M. Graña, and M. Cottrell. Basic Competitive Neural Networks as adaptive mechanisms for non-stationary colour quantisation. *Neural Computing & Applications*, 8(4):347–367, 1999.

[76] R.C. Gonzalez and R.E. Woods. *Digital image processing*. Addison-Wesley, Reading, MA, 1987.

[77] M. Graña. A brief review of Lattice Computing. *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, pages 1777–1781, June 2008.

[78] M. Graña, A. D'Anjou, A.I. González, F.X. Albizuri, and M. Cottrell. Competitive Stochastic Neural Networks for vector quantization of images. *Neurocomputing*, 7(2):187–195, 1995.

[79] M. Graña, I. Villaverde, J.O. Maldonado, and C. Hernandez. Two Lattice Computing approaches for the unsupervised segmentation of hyperspectral images. *Neurocomputing*, 72(10-12):2111–2120, 2009.

[80] M.D. Greicius, G. Srivastava, A.L. Reiss, and V. Menon. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc Natl Acad Sci U S A*, 101(13):4637–4642, 2004.

[81] R. Guillemaud and M. Brady. Estimating the bias field of MR images. *Medical Imaging, IEEE Transactions on*, 16(3):238–251, 1997.

[82] M.T. Hagan, H.B. Demuth, and M.H. Beale. *Neural Network design*. PWS Pub. Co., 1995.

[83] S. Haykin. *Neural Networks: a comprehensive foundation (2nd edition)*. Prentice Hall, 2 edition, 1998.

[84] T. Heskes. Energy functions for Self-Organizing Maps. *E. Oja, S. Kaskied (Eds.), Kohonen Maps, Elsevier, Amsterdam*, pages 303–315, 1999.

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

[86] C. Huang, B. Yan, H. Jiang, and D. Wang. Combining Voxel-Based Morphometry with Artifical Neural Network theory in the application research of diagnosing Alzheimer's disease. In *BMEI '08: Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics*, pages 250–254, Washington, DC, USA, 2008. IEEE Computer Society.

[87] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[88] K.M. Iftekharuddin, J. Zheng, M.A. Islam, and R.J. Ogg. Fractal-based brain tumor detection in multimodal MRI. *Applied Mathematics and Computation*, 207(1):23–41, 2009.

[89] C. Jean-François and A. Souloumiac. Blind beamforming for non gaussian signals. *IEE PROCEEDINGS-F*, 140:362–370, 1993.

[90] D.N. Kennedy, P.A. Filipek, and V.R. Caviness. Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging. *IEEE Trans. Med. Imaging*, 8(1):1–7, 1989.

[91] N. Keshava. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, 14(1):55–78, 2003.

[92] N. Keshava and J.F. Mustard. Spectral unmixing. *IEEE Signal Proc. Mag.*, 19(1):44–57, 2002.

[93] S. Kloppel. Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(3):681–689, 2008.

[94] T. Kohonen. *Self-Organization and Associative Memory*. Information Sciences. Springer, 3 sub edition, 1989.

[95] T. Kohonen. Learning Vector Quantization. *The handbook of brain theory and neural networks,MIT Press: Cambridge, MA, USA*, pages 537–540, 1998.

[96] T. Kohonen. *Self-Organizing Maps*. Springer Verlag, Berlin, third edition, 2000.

[97] R.K.-S. Kwan, A.C. Evans, and G.B. Pike. An extensible MRI Simulator for post-processing evaluation. In *VBC '96: Proceedings of the 4th International Conference on Visualization in Biomedical Computing*, pages 135–140, London, UK, 1996. Springer-Verlag.

[98] R.K.S. Kwan, A.C. Evans, and G.B. Pike. MRI simulation-based evaluation of image-processing and classification methods. *Medical Imaging, IEEE Transactions on*, 18(11):1085–1097, 1999.

[99] B.A. Landman, J.A.D. Farrell, C.K. Jones, S.A. Smith, J.L. Prince, and S. Mori. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *NeuroImage*, 36(4):1123–1138, 2007.

[100] O. Lange, A. Meyer Baese, M. Hurdal, and S. Foo. A comparison between neural and fuzzy cluster analysis techniques for functional MRI. *Biomedical Signal Processing and Control*, 1(3):243–252, 2006.

[101] Z. Lao, D. Shen, Z. Xue, B. Karacali, S.M. Resnick, and C. Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, 21(1):46–57, Jan 2004.

[102] C.L. Lawson and R.J. Hanson. *Solving least squares problems*. Englewood Cliffs : Prentice-Hall, 1974.

[103] J.A. Lee and M. Verleysen. Self-Organizing Maps with recursive neighborhood adaptation. *Neural Netw.*, 15(8-9):993–1003, 2002.

[104] S.Z. Li. Robustizing m-estimation using Deterministic Annealing. *Pattern Recognition*, 29(1):159–166, 1996.

[105] X. Li, L. Wang, and E. Sung. A study of AdaBoost with SVM based weak learners. In *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, volume 1, pages 196–201 vol. 1, 2005.

[106] Z. Liang, R.J. Jaszczak, and R.E. Coleman. Parameter estimation of finite mixtures using the EM algorithm and information criteria with application to medical image processing. *IEEE Trans. Nucl. Sci.*, 39:1126–1133, 1992.

[107] A.W.-C. Liew and H. Yan. An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation. *IEEE Trans. Med. Imaging*, 22(9):1063–75, Sep 2003.

[108] M. Lindquist. The statistical analysis of fMRI data. *Statistical Science*, 23(4):439–464, 2008.

[109] Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. Shenton, C.S. Carter, V.A. Stenger, S. Davis, H. Aizenstein, J.T. Becker, O.L. Lopez, and C.C. Meltzer. Discriminative MR image feature analysis for automatic schizophrenia and Alzheimer's disease classification. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 3216:393–401, 2004.

[110] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, and R.L. Buckner. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cognitive Neuroscience*, 19(9):1498–1507, 2007.

[111] J.L. Marroquin, B.C. Vemuri, S. Botello, and F. Calderon. An accurate and efficient bayesian method for automatic segmentation of brain MRI. In *Proceedings of the 7th European Conference on Computer Vision (ECCV) -Part IV*, pages 560–574, London, UK, 2002. Springer-Verlag.

[112] V. Megalooikonomou, C. Davatzikos, and E.H. Herskovits. A simulator for evaluating methods for the detection of lesion-deficit associations. *Human Brain Mapping*, 10(2):61–73, 2000.

[113] S. Mehta, T.J. Grabowski, Y. Trivedi, and H. Damasio. Evaluation of Voxel-Based Morphometry for focal lesion detection in individuals. *NeuroImage*, 20(3):1438–1454, November 2003.

[114] F.M. Miezin, L. Maccotta, J.M. Ollinger, S.E. Petersen, and R.L. Buckner. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage*, 11(6 Pt 1):735–59, June 2000. PMID: 10860799.

[115] M.L. Minsky and S.A. Papert. *Perceptrons*. Cambridge, MA: MIT Press, 1969.

[116] J.C. Morris. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*, 43(11):2412–2414, 1993.

[117] H.P. Muller, E. Kraft, A. Ludolph, and S.N. Erne. New methods in fMRI analysis. *Engineering in Medicine and Biology Magazine, IEEE*, 21(5):134–142, Sep/Oct 2002.

[118] N. Lazar. *The statistical analysis of functional MRI data (Statistics for Biology and Health)*. Springer, 1 edition, 2008.

[119] T.W. Nattkemper and A. Wismüller. Tumor feature visualization with unsupervised learning. *Med. Image Anal.*, 9(4):344–51, Aug 2005.

[120] N.J. Nilsson. *Learning Machines*. New York: MacGraw-Hill, 1965.

[121] D.B. Parker. Learning-Logic. *Invention Rep.*, pages 581–64, 1982.

[122] J.J. Pekar. A brief introduction to functional MRI. *Engineering in Medicine and Biology Magazine, IEEE*, 25(2):24–26, March-April 2006.

[123] D.L. Pham and J.L. Prince. Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Trans. Med. Imaging*, 18(9):737–52, Sep 1999.

[124] D.L. Pham and J.L. Prince. An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities. *Pattern Recogn. Lett.*, 20(1):57–68, 1999.

[125] B. Raducanu, M. Graña, and F.X. Albizuri. Morphological scale spaces and associative morphological memories: results on robustness and practical applications. *J. Math. Imaging Vis.*, 19(2):113–131, 2003.

[126] P.P. Raghu and B. Yegnanarayana. Supervised texture classification using a Probabilistic Neural Network and constraint satisfaction model. *IEEE Trans. Neural Networks*, 9:516–522, 1998.

[127] J.C. Rajapakse and F. Kruggel. Segmentation of MR images with intensity inhomogeneities. *Image Vision Comput.*, 16(3):165–180, 1998.

[128] J. Ramírez, M. Górriz, M. López, D. Salas-González, I. Álvarez, F. Segovia, and C.G. Puntonet. Early detection of the Alzheimer disease combining feature selection and kernel machines. *International Conference on Neural Information Processing of the Asia-Pacific Neural Network Assembly (ICONIP)*, 2:410–417, 2008.

[129] J. Ramírez, P. Yelamos, J.M. Górriz, and J.C. Segura. SVM-based speech endpoint detection using contextual speech features. *Electron. Lett.*, 42(7):426–428, 2006.

[130] G.X. Ritter, J.L. Diaz-de Leon, and P. Sussner. Morphological Bidirectional Associative Memories. *Neural Netw.*, 12(6):851–867, 1999.

[131] G.X. Ritter and P. Gader. Fixed points of lattice transforms and Lattice Associative Memories. *Hawkes, P. (Ed.), Advances in Imaging and Electron Physics*, 144:165–242, 2006.

[132] G.X. Ritter, P. Sussner, and J.L. Diza-de Leon. Morphological Associative Memories. *Neural Networks, IEEE Transactions on*, 9(2):281–293, Mar 1998.

[133] G.X. Ritter, G. Urcid, and L. Iancu. Reconstruction of patterns from noisy inputs using Morphological Associative Memories. *J. Math. Imaging and Vision*, 19(2):95–112, 2003.

[134] G.X. Ritter, G. Urcid, and M.S. Schmalz. Autonomous single-pass endmember approximation using Lattice Auto-Associative Memories. *Neurocomputing*, 72(10-12):2101–2110, 2009.

[135] H. Robbins and S. Monro. A stochastic approximation method. *Annals Math. Stat.*, 22:400–407, 1951.

[136] A. Rosenblatt. *Principles of neurodynamics*. New York: Spartan, 1959.

[137] E.H. Rubin, M. Storandt, J.P. Miller, D.A. Kinscherf, E.A. Grant, J.C. Morris, and L. Berg. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives of Neurology*, 55(3):395–401, March 1998. PMID: 9520014.

[138] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. *Parallel distributed processing: explorations in the microstructure of cognition*, 1: foundations:318–362, 1986.

[139] D.E. Rumelhart, J.L. McClelland, and the PDP Research Group. *Parallel distributed processing*, volume 1: Fundations. Cambridge, MA: MIT Press, 1986.

[140] D. Salas González, J. M. Górriz, J. Ramirez, M. López, I. Alvarez, F. Segovia, and C.G. Puntonet. Computer aided diagnosis of Alzheimer's disease using Support Vector Machine and Classification Trees. *International Conference on Neural Information Processing of the Asia-Pacific Neural Network Assembly (ICONIP)*, 2008.

[141] C.H. Salmond, J. Ashburner, F. Vargha-Khadem, A. Connelly, D.G. Gadian, and K.J. Friston. Distributional assumptions in Voxel-Based Morphometry. *NeuroImage*, 17(2):1027–1030, October 2002.

[142] O. Salvado, C. Hillenbrand, S. Zhang, and D.L. Wilson. Method to correct intensity inhomogeneity in MR images for atherosclerosis characterization. *IEEE Trans. Med. Imaging*, 25(5):539–52, May 2006.

[143] G.E. Sarty. *Computing brain activity maps from fMRI time-series images*. Cambridge University Press, 2007.

[144] A.S. Sato and K. Yamada. Generalized Learning Vector Quantization. *G. Tesauro, D. Touretzky, and T. Leen, editors, Advances in Neural Information Processing Systems, MIT Press*, 7:423–429, 1995.

[145] A.J. Saykin, L.A. Flashman, S.A. Frutiger, S.C. Johnson, A.C. Mamourian, C.H. Moritz, J.R. O'Jile, H.J. Riordan, R.B. Santulli, C.A. Smith, and J.B. Weaver. Neuroanatomic substrates of semantic memory impairment in Alzheimer's disease: patterns of functional MRI activation. *Journal of the International Neuropsychological Society*, 5:377–392, 1999.

[146] .E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Mach. Learn.*, 37(3):297–336, 1999.

[147] J. Sijbers, A.J. den Dekker, A. Van der Linden, T.M. Verhoye, and D. Van Dyck. Adaptive anisotropic noise filtering for magnitude MR data. *Magn. Reson. Imaging*, 17(10):1533–9, Dec 1999.

[148] M.Y. Siyal and L. Yu. An intelligent modified fuzzy C-means based algorithm for bias estimation and segmentation of brain MRI. *Pattern Recognition Letters*, 26(13):2052–2062, 2005.

[149] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998.

[150] S.M. Smith. Overview of fMRI analysis. *Br. J. Radiol.*, 77(Spec No 2):S167–S175, 2004.

[151] P. Somervuo and T. Kohonen. Self-Organizing Maps and Learning Vector Quantization for feature sequences. *Neural Process. Lett.*, 10(2):151–159, 1999.

[152] T. Song, M.M. Jamshidi, R.R. Lee, and M. Huang. A modified Probabilistic Neural Network for partial volume segmentation in brain MR image. *IEEE Trans. Neural Netw.*, 18(5):1424–32, Sep 2007.

[153] D. F. Specht. Probabilistic Neural Networks. *Neural Networks*, 3(1):109–118, 1990.

[154] D.F. Specht. Enhancements to the Probabilistic Neural Networks. *Proc. IEEE/INNS Joint Conf. Neural Networks (Baltimore, MD)*, pages 761–768, 1992.

[155] R.L. Streit and T.E. Luginbuhl. Maximum Likelihood training of Probabilistic Neural Network. *IEEE Trans. Neural Networks*, 5:764–783, 1994.

[156] S.C. Strother. Evaluating fMRI preprocessing pipelines. *Engineering in Medicine and Biology Magazine, IEEE*, 25(2):27–41, March-April 2006.

[157] M. Styner, C. Brechbühler, G. Székely, and G. Gerig. Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Trans. Med. Imaging*, 19(3):153–65, Mar 2000.

[158] J. Talairach and P. Tournoux. *Co-planar stereotaxic atlas of the human brain*. Thieme New York, 1988.

[159] D. Tao, X. Tang, X. Li, and X. Wu. Asymmetric bagging and random subspace for Support Vector Machines-based relevance feedback in image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(7):1088–1099, 2006.

[160] Q. Tianbai and L. Minglu. Multispectral MR images segmentation using SOM network. In *Proceedings of the The Fourth International Conference on Computer and Information Technology (CIT)*, pages 155–158, Washington, DC, USA, 2004. IEEE Computer Society.

[161] M.E. Tipping and D. Lowe. Shadow targets: a novel algorithm for topographic projections by radial basis functions. *Neurocomputing*, 19(1-3):211–222, 1998.

[162] H.G.C. Traven. A Neural-Network approach to statistical pattern classification by semiparametric estimation of a probability density funcitons. *IEEE Trans. Neural Networks*, 2:366–377, 1991.

[163] V. Vapnik. *Statistical learning theory*. Wiley-Interscience, 1998.

[164] C. Varini, A. Degenhard, and T.W. Nattkemper. Visual exploratory analysis of DCE-MRI data in breast cancer by dimensional data reduction: a comparative study. *Biomedical Signal Processing and Control*, 1(1):56–63, 2006.

[165] P. Vemuri, J.L. Gunter, M.L. Senjem, J.L. Whitwell, K. Kantarci, D.S. Knopman, B.F. Boeve, R.C. Petersen, and C.R. Jack. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *NeuroImage*, 39(3):1186–1197, Feb 2008.

[166] T. Villmann, R. Der, M. Herrmann, and T .M. Martinetz. Topology preservation in self-organizing feature maps: exact definition and measurement. *IEEE Trans. Neural Netw.*, 8(2):256–66, 1997.

[167] U. Vovk, F. Pernus, and B. Likar. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans Med Imaging*, 26(3):405–21, Mar 2007.

[168] W.M. Wells, W.L. Grimson, R. Kikinis, and F.A. Jolesz. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imaging*, 15(4):429–42, 1996.

[169] P. Werbos. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. Ph.D. thesis, Harvard Univ., Cambridge, MA, 1974.

[170] H. White. Learning in Artificial Neural Networks: a statistical perspective. *Neural Computation*, 1:425–464, 1989.

[171] R. Winter and B. Widrow. MADALINE RULE II: a training rule for Neural Networks. *Proc. IEEE/INNS Joint Conf. Neural Networks (San Diego, CA)*, 1:401–408, 1988.

[172] K. J. Worsley, A.C. Evans, S. Marrett, and P. Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12(6):900–918, November 1992. PMID: 1400644.

[173] K.J. Worsley and K.J. Friston. Analysis of fMRI time-series revisited - again. *NeuroImage*, 2(3):173–181, 1995.

[174] M.X.H. Yan and J.S. Karp. An Adaptive Bayesian approach to three dimensional MR brain segmentation. *Proceedings of Information Processing and Medical Imaging (IPMI)*, pages 201–213, 1995.

[175] A. Younis, M. Ibrahim, M. Kabuka, and N. John. An Artificial Immune-Activated Neural Network applied to brain 3D MRI segmentation. *J. Digit. Imaging*, 21 Suppl 1:S69–88, Oct 2008.

[176] A. Zaknich. A vector quantization reduction method for the Probabilistic Neural Network. *Proc. IEEE/INNS Joint Conf. Neural Networks (Piscataway, NJ)*, pages 1117–1120, 1997.

[177] J. Zhang, B. Yan, X. Huang, P. Yang, and C. Huang. The diagnosis of Alzheimer's disease based on Voxel-Based Morphometry and Support Vector Machine. In *Proceedings of the 2008 Fourth International Conference on Natural Computation (ICNC)*, pages 197–201, Washington, DC, USA, 2008. IEEE Computer Society.

[178] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a Hidden Markov Random Field model and the Expectation-Maximization algorithm. *IEEE Trans. Med. Imaging*, 20(1):45–57, Jan 2001.