KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK (ESAT),
AFDELING PSI
Kasteelpark Arenberg 10, 3001 Leuven-Heverlee (Belgium)

FACULTEIT GENEESKUNDE
DEPARTEMENT MORFOLOGIE EN MEDISCHE
BEELDVORMING, AFDELING RADIOLOGIE
Herestraat 49, 3000 Leuven (Belgium)

Quantitative Analysis of Signal Abnormalities in MR Imaging

for Multiple Sclerosis and Creutzfeldt-Jakob Disease

Promotoren:
prof. Dr. ir. D. Vandermeulen
prof. Dr. ir. P. Suetens

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen
door

Koen VAN LEEMPUT

18 mei 2001

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK (ESAT),
AFDELING PSI
Kasteelpark Arenberg 10, 3001 Leuven-Heverlee (Belgium)

FACULTEIT GENEESKUNDE
DEPARTEMENT MORFOLOGIE EN MEDISCHE
BEELDVORMING, AFDELING RADIOLOGIE
Herestraat 49, 3000 Leuven (Belgium)

Quantitative Analysis of Signal Abnormalities in MR Imaging

for Multiple Sclerosis and Creutzfeldt-Jakob Disease

Leden van de jury:
prof. Dr. ir. J. Berlamont (voorzitter)
prof. Dr. ir. D. Vandermeulen (promotor)
prof. Dr. ir. P. Suetens (promotor)
prof. Dr. ir. F. Maes
prof. Dr. ir. S. Van Huffel
prof. Dr. ir. J. Nuyts
prof. Dr. P. Demaerel
prof. Dr. W. Van Paesschen
prof. Dr. ir. G. Gerig (UNC Chapel Hill)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen
door

Koen VAN LEEMPUT

U.D.C. 615.849

18 mei 2001

ISBN 90-5682-299-3
D/2001/7515/13

Theory is important

– at least in theory

(K. Martin)

Acknowledgements

This thesis is the result of many invisible hands helping me. Special thanks go to my supervisor and mentor Dirk Vandermeulen. His unique way of steering my research by seemingly accidentally leaving papers on my desk and by giving subtle hints at the right moment, kept me on steady course. I am convinced that he will recognize many of his original ideas in one form or another while browsing through this thesis. My gratitude also to my other supervisor, Paul Suetens, for having given me the opportunity to work in his internationally recognized group and for taking care of everything that could have diverted me from my research.

I am deeply appreciative to the second author of my publications, Frederik Maes. His willingness to patiently read and re-read my drafts and his ability to mercilessly filter out weak points have greatly improved the quality of my papers. Working and discussing with him has been a great pleasure and a very instructive experience.

During the often memorable BIOMORPH and QAMRIC meetings, I have had the pleasure to get to know some of the leading researchers in the field. Amongst them, I would like to thank Alan Colchester, Nicholas Ayache, Mike Brady and Guido Gerig for their interest and for their many useful comments. Special thanks also to Eric Bardinet and Nicholas Ayache for my most enjoyable stay with the Epidaure group at INRIA Sophia-Antipolis last September.

I am grateful to my colleagues and to all the people somehow involved in the Leuven group for the refreshing and open atmosphere (figuratively speaking, that is). Special thanks go to all those who have ever helped me with one of my many computer problems. Since computer-related topics are not my strongest side, I am afraid that includes almost everyone.

I thank the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT) for its financial support.

A special word of thank goes to the people from what is now called the medical imaging research group in Oulu, Finland. In particular, I would like to thank Lasse Jyrkinen and Olli Silvén for helping me on my first steps in the field of medical imaging. Not only do they lie directly on the basis of this PhD by infecting me with the research virus at the time of my MSc thesis, their idea of having me come over during the summer of 1996 also changed my life in many other ways as well.

Addictive and fascinating as the research part of this PhD was, the writing was tedious and difficult. I would therefore like to thank my family for their support: my caring parents; my brother Dirk, who has been jokingly reminding me to mention his name in this acknowledgement ever since he heard that I had started writing up; his girlfriend Yvonne; and especially my girlfriend Elena. Elena, your enthusiasm, your many brilliant ideas, your patient listening to my frustrations, your encouraging words and your strong belief in the formability of life, make you the ghostwriter of this thesis. Thank you for always being there for me. Thank you for painting my life in the most beautiful colors.

Leuven, May 4, 2001

Abstract

Quantitative analysis of signal abnormalities in MR imaging for multiple sclerosis and Creutzfeldt-Jakob disease

In multiple sclerosis and Creutzfeldt-Jakob disease, pathological changes in the brain result in deviations in signal intensity on MR images. Quantitative analysis of these changes and their correlation with clinical findings provides important information for *in vivo* diagnosis, therapy planning and follow-up as well as for biomedical research. This thesis presents a framework for fully automated, objective and reproducible quantification of pathology-related signal abnormalities in large amounts of multi-spectral MR data. The overall strategy adopted was to build statistical models for normal brain MR images, including explicit models for MR field inhomogeneities, tissue-specific intensity distributions, spatial clustering of tissue types, and the partial volume effect. Signal abnormalities were detected as model outliers, i.e. voxels that could not be well explained by the model. Throughout this thesis, special attention was paid to automatically estimate all model parameters from the data itself, to eliminate subjective manual tuning and training.

Kwantitatieve analyse van signaalabnormaliteiten in MR-beeldvorming voor multiple sclerose en Creutzfeldt-Jakob

Bij multiple sclerose en de ziekte van Creutzfeldt-Jakob resulteren pathologische veranderingen in de hersenen in afwijkingen in de signaalintensiteit op MR-beelden. Kwantitatieve analyse van deze veranderingen en hun correlatie met klinische bevindingen levert belangrijke informatie op voor zowel *in vivo* diagnose, therapieplanning en opvolging als voor biomedisch onderzoek. Deze thesis stelt een raamwerk voor voor volledig automatische, objectieve en reproduceerbare kwantificatie van pathologische signaalafwijkingen in grote hoeveelheden multi-spectrale MR-data. De algemene strategie die werd gevolgd, bestond uit het construeren van statistische modellen voor MR-beelden van normale hersenen, met inbegrip van expliciete modellen voor MR-veldinhomogeniteiten, weefselspecifieke intensiteitsverdelingen, spatiale clustering van weefseltypes, en het partieel-volume-effect. Signaalabnormaliteiten werden gedetecteerd als voxels die niet goed door het model konden worden verklaard. Doorheen deze thesis werd er bijzondere aandacht geschonken aan het automatisch schatten van alle modelparameters uit de data zelf, om subjectieve manuele interventie uit te sluiten.

Contents

List of acronyms and mathematical symbols

2-D 2-dimensional
3-D 3-dimensional
BSE ..... Bovine Spongiform Encephalopathy
CJD ..... Creutzfeldt-Jakob disease
CSF ..... Cerebrospinal Fluid
EDSS ..... Extended Disability Status Scale
EM ..... Expectation-Maximization
EU ..... European Union
FLAIR ..... Fluid Attenuated Inversion Recovery
ML ..... Maximum Likelihood
MR ..... Magnetic Resonance
MRF ..... Markov Random Field
MRI ..... Magnetic Resonance Imaging
MS ..... Multiple Sclerosis
PCA ..... Principal Component Analysis
PD ..... Proton Density
PDM ..... Point Distribution Model
PrP ..... Prion Protein
PV ..... Partial Volume
TLL ..... Total Lesion Load
$E_X[\,\cdot \mid Y\,]$ ..... expected value over X given Y
$G_{\mathbf{\Sigma}}(\cdot)$ ..... zero-mean normal distribution with covariance $\mathbf{\Sigma}$

Kwantitatieve analyse van signaalabnormaliteiten in MR-beeldvorming voor multiple sclerose en Creutzfeldt-Jakob (in Dutch)

Inleiding

Verscheidene neuropathologieën van het centrale zenuwstelsel zoals multiple sclerose (MS), schizofrenie, epilepsie, Alzheimer en de ziekte van Creutzfeldt-Jakob (CJD), hebben te maken met morfologische en/of structurele veranderingen in de hersenen. Het is van groot belang om zulke aandoeningen te kunnen bestuderen door op een objectieve en directe manier deze veranderingen te meten i.p.v. zich volledig te moeten baseren op de klinische symptomen. Om de diagnose van de (nieuwe) variante van de ziekte van Creutzfeldt-Jakob te ondersteunen, zou het bijvoorbeeld nuttig zijn om veranderingen in de hersenen te kunnen kwantificeren, aangezien er op dit moment geen standaard klinische diagnostische criteria voor deze ziekte bestaan. Voor schizofrenie zou het meten van de asymmetrie in volume en vorm tussen linker versus rechter hersenhelftstructuren kunnen helpen om de pathogenese van de ziekte beter te begrijpen. In grote klinische trials, waar wordt geëvalueerd of een nieuwe therapie de evolutie van een ziekte al dan niet gunstig beïnvloedt, is het extreem duur om direct een reductie van de klinische symptomen te proberen aantonen voor zulk een trage en variabele ziekte als MS. Kwantificatie van morfologische veranderingen in de hersenen en correlatie van deze veranderingen met klinische bevindingen zou alternatieve eindpunten kunnen opleveren die vlugger veranderen en betrouwbaarder te meten zijn. Zulke metingen kunnen ook belangrijke informatie leveren voor de therapieplanning en de opvolging van individuele patiënten.

Hedendaagse medische beelvormingsmodaliteiten maken het mogelijk om op een niet-invasieve manier kwantitatieve informatie over de menselijke anatomie in drie dimensies te bekomen. Wegens zijn hoge spatiale resolutie en zijn inherent goed contrast in zachte weefsels is magnetische-resonantiebeeldvorming (MRI) superieur aan andere beeldvormingstechnieken om aandoeningen van het centrale zenuwstelsel te bestuderen. Traditioneel is de interpretatie van MR-beelden gebaseerd op dwarsdoorsneden afgebeeld op film, wat alleen subjectieve kwalitatieve en beschrijvende resultaten toelaat. Het digitale karakter van MRI biedt echter de mogelijkheid om objectieve en nauwkeurige kwantitatieve informatie uit de beelden te extraheren. Zulke kwantitatieve metingen kunnen vergeleken worden over de tijd heen voor het opvolgen van morfologische veranderingen in één individu (bijvoorbeeld om het verloop van MS te evalueren), tussen individuen, of tussen populaties (bijvoorbeeld schizofrenen versus controlepatiënten; groepen die een geneesmiddel toegediend krijgen versus placebogroepen).

Manuele analyse van de beelden door een getrainde menselijke expert is een omslachtige en uiterst moeizame taak aangezien de structuren van belang complexe randen vertonen in 3-D en dikwijls geen duidelijk zichtbare anatomische grenzen hebben. In klinische trials is het aantal MR-beelden zo groot dat manuele analyse door menselijke experts extreem tijdrovend is. Bovendien is het onduidelijk hoe

een menselijke waarnemer de informatie uit de verschillende kanalen combineert wanneer multi-spectrale MR-beelden onderzocht worden. Verder bemoeilijkt de variabiliteit tussen aflijningen van verschillende waarnemers en tussen herhaalde aflijningen door eenzelfde waarnemer de analyse van de resultaten.

Daarom is er nood aan volledig automatische methoden voor kwantificatie van MR-beelden van de hersenen die op een reproduceerbare manier grote hoeveelheden multi-spectrale MR-data kunnen analyseren. Deze thesis concentreert zich op de automatische kwantificatie van pathologie-gerelateerde *intensiteits*abnormaliteiten in MR-beelden van patiënten die lijden aan MS of CJD. De ontwikkelde technieken zijn algemeen genoeg om ook toepasbaar te zijn op andere neurologische aandoeningen met MRI-signaalveranderingen, zoals bijvoorbeeld leukomalacie, maar de analyse van abnormaliteiten in *vorm*, zoals nodig in bijvoorbeeld epilepsie en schizofrenie, valt buiten het bestek van dit werk.

Multiple sclerose

Multiple sclerose is een relatief vaak voorkomende ziekte bij jonge volwassenen die vooral de witte stof van het centrale zenuwstelsel aantast. Hoewel 10 procent van de patiënten het goed stelt gedurende langer dan 20 jaar, heeft 50 procent hulp nodig bij het stappen binnen de 15 jaar, en in zeldzame gevallen sterven patiënten binnen enkele maanden na het optreden van de eerste symptomen. Het relatief aantal MS-gevallen varieert aanzienlijk over de wereld, en is het hoogst in noord-Europa, zuid-Australië en het middengedeelte van noord-Amerika. Hoewel de oorzaak van de ziekte onbekend is, gelooft men dat omgevings- en genetische factoren een rol spelen.

Relapsing-remitting MS is het type dat voorkomt in 80 procent van de gevallen, en treft dubbel zo vaak vrouwen als mannen. Typisch verslechteren de symptomen over een periode van verscheidene dagen, om vervolgens te stabiliseren en dikwijls opnieuw te verbeteren. Blijvende tekenen van disfunctie van het centrale zenuwstelsel kunnen zich echter ontwikkelen na elke terugval, en de ziekte kan zich ook tussen zulke terugvallen verder ontwikkelen. Primair-progressieve MS komt voor in 20 procent van de patiënten, en wordt gekenmerkt door een gradueel progressief klinisch verloop en door een gelijkaardig aantal gevallen bij mannen en vrouwen.

MRI is gekend als de bij uitstek beste paraklinische test voor MS, en toont letsels in de beelden in 95 procent van de patiënten. Diagnose gaat vandaag praktisch altijd gepaard met een bevestigende MR-studie. Het risico op terugval en de graad van handicap kunnen tot op zekere hoogte voorspeld worden vanuit de bevindingen op MR-beelden van de hersenen tijdens de eerste klinische episode. Verder wordt MRI steeds meer gebruikt om traditionele neurologische schalen voor de graad van handicap te ondersteunen bij het opvolgen van de progressie van MS en bij het evalueren van het effect van een nieuw geneesmiddel. Eén van de redenen hiervoor is dat MR-metingen veel gevoeliger zijn en duidelijk een belangrijk aspect van het onderliggende pathologische proces blootleggen. Het groeiend gebruik van MRI in MS maakt de ontwikkeling van automatische methoden voor het bekomen van nauwkeurige, reproduceerbare metingen van de letsels in grote hoeveelheden beelddata

noodzakelijk.

Creutzfeldt-Jakob

De ziekte van Creutzfeldt-Jakob is de meest voorkomende menselijke vorm van zogenaamde overdraagbare spongiforme encephalopathieën of prionziekten, een groep van neurodegeneratieve aandoeningen die zowel mensen als dieren treffen. Hoewel gevallen van CJD zeldzaam zijn, geniet de ziekte een grote belangstelling wegens haar eigenaardig causatief agens met nieuwe vormen van vermenigvuldiging en overdraging, nl. een abnormale isovorm van een normaal proteïne. Bovendien is er in de media uitgebreid aandacht geschonken aan het verschijnen van BSE bij runderen ("gekke-koeien-ziekte"), zijn snelle evolutie tot een epidemie in Groot-Brittannië, zijn overdraging naar de mens, en zijn mogelijk ernstige bedreiging voor de volksgezondheid.

In 85 tot 90 procent van de gevallen van CJD gaat het om zogenaamde sporadische CJD, een aandoening met ongekende oorzaak die voornamelijk ouderen treft en die de dood veroorzaakt binnen gemiddeld 5 maanden. Tussen 10 en 15 procent van de CJD-patiënten lijdt aan erfelijk overgedragen CJD, en in zeldzame gevallen wordt de aandoening opgelopen als gevolg van kannibalisme, besmetting via chirurgische instrumenten, weefseldonatie of injectie met menselijk groeihormoon. In 1995 werden drie patiënten met een nieuwe variant van CJD gesignaleerd in Groot-Brittannië, en dat aantal is ondertussen opgelopen tot 85 bevestigde gevallen. De patiënten zijn jonger dan deze getroffen door de meer vertrouwde vormen van de aandoening, en er wordt algemeen aangenomen dat deze vorm van CJD de menselijk variant is van BSE.

Tot op de dag van vandaag kan een definitieve diagnose van CJD alleen gemaakt worden door hersenbiopsie of necroscopie, of door biopsie van de amandelen voor de nieuwe variant van CJD. Aangezien men subtiele intensiteitsabnormaliteiten heeft opgemerkt op MR-hersenbeelden van CJD-patiënten, zou MRI een potentiële niet-invasieve test voor vroege diagnose van CJD kunnen zijn. Om dit te kunnen onderzoeken, is er nood aan accurate, automatische methoden om MR-intensiteiten in specifieke hersengebieden en intensiteitsverschillen tussen zulke gebieden te kwantificeren en te vergelijken tussen CJD-groepen en controlegroepen.

Algemene oplossingsstrategie

Automatische kwantitatieve analyse van intensiteitsabnormaliteiten in MR-beelden van de hersenen vormt een zeer moeilijk probleem. De complexe 3-D structuur van de hersenen en ambiguïteiten in de beelden veroorzaakt door het beeldvormingsproces, zoals beperkte resolutie, partieel-volume-effecten, ruis, gebrek aan contrast, intensiteitsinhomogeniteiten en andere artefacten, maken het zelfs voor menselijke experts moeilijk om de beelden te interpreteren. Om deze problemen het hoofd te kunnen bieden, werden in deze thesis modellen gebruikt die voorkennis beschrijven over het verwachte uitzicht van de hersenen in MR-beelden. De algemene strategie

die werd gevolgd, bestond uit het construeren van steeds nauwkeurigere statistische modellen voor MR-beelden van normale hersenen, en uit het detecteren van signaalabnormaliteiten als voxels die niet goed door zulke modellen kunnen worden verklaard.

Een ongebruikelijk aspect van het werk beschreven in deze thesis is dat alle modelparameters telkens worden geschat uit de data zelf, vertrekkend van een initialisatie die wordt verkregen zonder gebruikersinterventie. Dit zorgt ervoor dat de modellen zich volledig automatisch op elke individuele scan hertrainen, wat de methode toelaat om beelden te analyseren met een niet eerder geziene MR-pulssequentie en voxelgrootte. Op deze manier wordt elke vorm van subjectieve manuele interventie vermeden, zodat de resultaten volledig objectief en reproduceerbaar zijn.

Automatische biasveldcorrectie

Een belangrijk probleem voor de automatische analyse van MR-beelden is de aanwezigheid van een zacht variërende intensiteitsinhomogeniteit of biasveld, wat resulteert in een niet-uniforme intensiteit van de weefsels doorheen het beeld. Deze bias is inherent aan MR-beeldvorming en wordt veroorzaakt door technische beperkingen en patiënt-geïnduceerde elektrodynamische interacties. Hoewel het niet altijd zichtbaar is voor het menselijk oog, kan het ernstige misclassificaties veroorzaken wanneer intensiteits-gebaseerde segmentatietechnieken worden gebruikt. Correctie van de MR-intensiteiten voor biasveldinhomogeniteiten is daarom een noodzakelijke voorwaarde voor robuuste automatische segmentatie.

Gebaseerd op de bestaande literatuur werd een model-gebaseerde methode ontwikkeld voor volledig automatische biasveldcorrectie van MR-beelden van de hersenen. Het weefseltype van elke voxel werd statistisch onafhankelijk verondersteld van het weefseltype van alle andere voxels. De MR-beelden werden gemodelleerd door weefselafhankelijke normaal-verdeelde ruis toe te kennen aan elke voxel, en er vervolgens een polynomiaal biasveldmodel bij op te tellen.

Segmentatie tracht het onderliggende weefseltype in elke voxel te reconstrueren gebaseerd op de waargenomen MR-beelden. Hiervoor moeten de modelparameters, bestaande uit de gemiddelde intensiteit van elke weefeltype en de variantie errond enerzijds, en de veeltermcoëfficiënten van het biasveldmodel anderzijds, gekend zijn. De schatting van deze parameters wordt echter op haar beurt geholpen door kennis van de segmentatie. Intuïtief kunnen zowel de segmentatie als de modelparameters tegelijkertijd geschat worden door de segmentatie af te wisselen met een herschatting van de parameters. Het EM-algoritme formaliseert deze intuïtieve benadering door iteratief het ongekende weefseltype in elke voxel te schatten gebaseerd op de huidige parameterschatting en de waargenomen MR-intensiteiten, en vervolgens de overeenkomstige modelparameters te berekenen. Dit leidt tot een iteratief algoritme dat telkens een classificatie van de voxels afwisselt met een schatting van de weefselspecifieke normaalverdelingen en een schatting van het biasveld. Het biasveld wordt geschat als een gewogen kleinste-kwadratenbenadering van het verschilbeeld tussen een voorspeld beeld zonder biasveld en het waargenomen MR-beeld, met gewichten omgekeerd evenredig met de variantie van het weefseltype waartoe elke voxel

geclassificeerd is. Het biasveld wordt daarom hoofdzakelijk berekend uit voxels die tot weefsels behoren met een smalle intensiteitsverdeling, zoals witte en grijze stof, waar het op een betrouwbare manier kan geschat worden, en geëxtrapoleerd naar regio's waar zo een schatting slecht geconditioneerd is (CSF, niet-hersenweefsel).

In eerdere methoden werden weefsel-specifieke intensiteitsmodellen geconstrueerd door manueel representatieve punten te selecteren voor elk van de in rekening gebrachte weefsels, wat moet herhaald worden voor elke set van gelijkaardige MR-beelden. Onze methode hertraint zulke modellen echter volledig automatisch op elke individuele dataset, vertrekkend van een initiële schatting van de classificatie die wordt afgeleid uit een digitale hersenatlas. Deze atlas bevat informatie over de a priori verwachte verdeling van witte stof, grijze stof en CSF, en wordt eerst automatisch in spatiale overeenstemming gebracht met de beelden gebruik makend van een affiene multi-modale registratietechniek gebaseerd op maximalisatie van mutuele informatie. Deze manier van werken laat volledige automatisatie van de methode toe zonder gebruikersinteractie, wat meer objectieve en reproduceerbare resultaten oplevert.

Modelleren van spatiale context

Als gevolg van de veronderstelling dat het weefeltype van elke voxel onafhankelijk is van het weefseltype van alle andere voxels, worden de voxels alleen geclassificeerd op basis van hun intensiteit. Dit levert aanvaardbare segmentatieresultaten op zolang de verschillende weefsels een duidelijk onderscheidbare intensiteitsverdeling hebben. Dit is echter niet altijd het geval voor MR-beelden van het hoofd, vooral niet wanneer er slechts een uni-spectrale MR-opname voorhanden is. Een typische probleem is bijvoorbeeld dat voxels in weefsels rondom de hersenen vaak een MR-intensiteit vertonen die zeer gelijkaardig is aan die van hersenweefsel. Dit kan resulteren in foutieve classificaties van kleine regio's rond de hersenen als grijze stof of witte stof.

Om zulke misclassificaties te ontmoedigen, werd het model uitgebreid door algemene spatiale en anatomische beperkingen in rekening te brengen tijdens de classificatie, gebruik makend van een Markov random veld (MRF) model. Dit MRF werd ontworpen om het onderscheid tussen hersen- en niet-hersenweefsel te vergemakkelijken en tegelijkertijd de gedetailleerde randen tussen de verschillende weefselklassen binnenin de hersenen te vrijwaren. Aangezien de exacte berekening van de classificatie in praktijk niet mogelijk is met het MRF-model, werd een zogenaamde *mean field*-benadering gebruikt die gebaseerd is op de veronderstelling dat de invloed van alle andere voxels op de classificatie van elke voxel kan benaderd worden door hun classificatie uit de vorige iteratie. Het resultaat is dat de a priori waarschijnlijkheid dat een voxel tot een bepaalde weefselklasse behoort, afhankelijk wordt van de classificatie van zijn buurvoxels.

Het resulterende EM-algoritme wisselt iteratief 4 stappen af: classificatie van de voxels, schatting van de normaalverdelingen, schatting van het biasveld, en schatting van de MRF-parameters. De methode werd gevalideerd op gesimuleerde MR-beelden van de hersenen, en op echte MR-data door de automatische segmentaties

te vergelijken met manuele aflijningen uitgevoerd door een menselijke expert. In beide gevallen waren fouten in de automatische segmentaties voornamelijk te wijten aan het verkeerd classificeren van de rand tussen de witte en de grijze stof, waar zogenaamde partieel-volume (PV) voxels niet exclusief tot witte of tot grijze stof behoren, maar een echte mengeling van beide vormen.

Modeloutliers en robuuste parameterschatting

Tot dusver werd er alleen aandacht geschonken aan de segmentatie van MR-beelden van normale hersenen. Om MS-letsels of CJD-gerelateerde signaalabnormaliteiten in de beelden te kunnen kwantificeren, werd de methode verder uitgebreid. Het is moeilijk om een expliciet model voor de pathologische weefsels te vinden omdat hun intensiteit in MR-beelden sterk varieert en omdat niet elke individueel beeld genoeg pathologie bevat om er de modelparameters uit te kunnen schatten.

Deze problemen werden omzeild door letsels te detecteren als voxels die niet goed door het statistisch model voor normale MR-hersenbeelden kunnen worden verklaard. Gebaseerd op het principe van de M-schatters uit het domein van de robuuste statistiek, werden weefsel-specifieke gewichten aan de voxels toegekend die reflecteren hoe typisch de voxels zijn voor elke weefselklasse. Deze gewichten zijn afhankelijk van een a priori gekozen statistisch significantieniveau dat definieert wanneer het geloof dat een voxel een outlier is, groter wordt dan het geloof dat deze voxel tot normaal weefsel behoort. Toevoeging van deze gewichten resulteert in een robuust EM-algoritme dat iteratief afwisselt tussen een statistische classificatie van de voxels in normale weefeltypes, inschatting van het geloof voor elke voxel om tot een letsel te behoren, en, enkel gebaseerd op wat wordt beschouwd als normaal weefsel, herschatting van de modelparameters.

Toepassing op multiple sclerose

Dit schema voor outlierdetectie werd toegepast voor volledig automatische segmentatie van MS-letsels uit multi-spectrale MR-hersenbeelden. Outliervoxels komen ook buiten MS-letsels voor, typisch in PV-voxels die, in tegenstelling tot de veronderstellingen die worden gemaakt in het statistisch model, niet tot één enkel weefseltype behoren maar bestaan uit een mengeling van verschillende weefsels tegelijkertijd. Om te verhinderen dat deze voxels, die tot normaal weefsel behoren, als MS-letsel worden geclassificeerd, werden bijkomende intensiteitsbeperkingen ingevoerd, tezamen met een MRF dat classificatie van voxels buiten de witte stof als MS-letsels ontmoedigt.

De methode werkt volledig automatisch, met slechts één enkele parameter die experimenteel moet bepaald worden nl. een statistisch significantieniveau. Dit contrasteert met het half-automatische karakter van de meeste bestaande methoden voor MS-letselsegmentatie, waarbij typisch de gebruiker automatisch berekende letsels moet accepteren of verwerpen, of manueel regio's van zuivere weefseltypes moet aanduiden om een classificator te trainen. Zulke methoden zijn onpraktisch wanneer

honderden beelden moeten geanalyseerd worden in een klinische trial, en de variabiliteit die gepaard gaat met manuele aflijningen blijft bestaan. Sommige bestaande technieken zijn wel volledig automatisch, maar gebruiken een vaste classificator die slechts één keer getraind wordt en vervolgens wordt gebruikt om honderden beelden te analyseren. In klinische trials kunnen echter intensiteitsverschillen tussen overeenkomstige weefsels in verschillende beelden niet uitgesloten worden, niet alleen vanwege veranderingen in de hardware van MR-scanners over de tijd heen, maar ook omdat mogelijk verschillende scanners gebruikt worden in een multi-center trial. In tegenstelling tot deze methoden hertraint ons algoritme de classificator op elke individuele scan, waardoor het zich aanpast aan zulke contrastvariaties. Bovendien worden ook de MRF-parameters automatisch herschat op elk individueel beeld, waardoor wordt vermeden dat de traditioneel veel gebruikte morfologische operatoren, connectiviteitsregels, minimumgrootte enz. om vals-positieve voxels te onderdrukken, opnieuw experimenteel moeten ingesteld worden telkens wanneer beelden met een andere voxelgrootte worden gesegmenteerd.

De methode werd gevalideerd door het totale letselvolume (TLL) in manuele segmentaties uitgevoerd door een menselijke expert te vergelijken met die van de automatische techniek op 20 beelden, voor een groot aantal verschillende waarden van het statistische significantieniveau. Hoewel de gemiddelde TLL geproduceerd door de automatische methode varieerde van slechts 25 % tot 150% van de gemiddelde TLL geschat door de expert, bleven de automatische metingen altijd evenredig veranderen met de manuele metingen, met hoge correlatiecoëfficiënten tussen 0.96 en 0.98. Daarom is de exacte keuze van het statistisch significantieniveau vrij onbelangrijk in klinische trials, die veranderingen in TLL proberen te meten. De keuze is echter veel kritischer wanneer het doel erin bestaat om de basismechanismen van MS of tijdscorrelaties van letselgroepen in MS-tijdseries te onderzoeken, aangezien deze toepassingen vereisen dat de letsels ook spatiaal correct gedetecteerd worden. In het algemeen geldt dat hoe hoger de resolutie is en hoe beter het contrast tussen letsels en normaal weefsel in de beelden, hoe gemakkelijker MS-letsels gedetecteerd worden door de automatische methode en hoe hoger het significantieniveau moet worden gekozen.

Toepassing op Creutzfeldt-Jakob

Toepassing van de techniek voor outlierdetectie bleek niet altijd succesvol voor de automatische kwantificatie van de vaak subtiele MR-intensiteitsafwijkingen in CJD. Hoewel er in een aantal gevallen abnormale intensiteiten werden gedetecteerd in structuren in de diepe grijze stof, werd ook een belangrijke hoeveelheid PV-voxels als outlier gesegmenteerd, vooral langsheen de rand tussen grijze stof en CSF. Voor de segmentatie van MS-letsels werden zulke vals-positieven onderdrukt door een MRF dat de detectie van MS-letsels buiten de witte stof ontmoedigt. Het bleek echter niet mogelijk om met een dergelijke techniek PV-voxels te onderscheiden van echt pathologische voxels in CJD, waar de intensiteitsafwijkingen veel subtieler zijn. Er werd daarom besloten dat het partieel-volume-effect expliciet moet worden opgenomen in het model, enerzijds om te voorkomen dat PV-voxels worden gede-

tecteerd als modeloutliers, en anderzijds om een realistischer model te bekomen dat nauwer aansluit met de data, zodat subtiele intensiteitsafwijkingen gemakkelijker gedetecteerd kunnen worden.

Partieel-volumesegmentatie

Het statistisch model werd daarom verder uitgebreid met een downsampling-stap waarin een aantal voxels in het oorspronkelijke beeldrooster samen de intensiteit van elke voxel in het resulterende beeldrooster bepalen. Dit veroorzaakt partieel volume in voxels waarvan niet alle onderliggende voxels tot het zelfde weefseltype behoren. Een EM-algoritme voor het schatten van de modelparameters werd afgeleid dat het algoritme dat in de voorgaande delen werd gebruikt, omvat als een speciaal geval. Er werden drie verschillende a priori spatiale modellen onderzocht: 1) de kans op een bepaalde mengverhouding tussen verschillende weefsels is onafhankelijk van de plaats in het beeld; 2) een gelijkaardig model maar waarbij alle mengverhoudingen even waarschijnlijk zijn wanneer twee weefsels mengen in een voxel; en 3) een MRF-model. Dit resulteert in een algemeen raamwerk voor partieel-volumesegmentatie dat het mogelijk maakt om weefsel-specifieke intensiteitsverdelingen te schatten geleid door spatiale informatie, terwijl de beeldvoxels tegelijkertijd geclassificeerd worden. Er werd aangetoond dat de methode bestaande technieken omvat en uitbreidt, waarbij een aantal gebreken en foutieve veronderstellingen in de bestaande methoden geïdentificeerd en verbeterd werden. Er werd bijvoorbeeld aangetoond dat een veelgebruikt MRF-model, waarbij simpelweg wordt opgelegd dat naburige voxels gelijkaardige mengverhoudingen tussen verschillende weefsels moeten hebben, tot foutieve resultaten leidt, en dat de gebruikelijke veronderstelling dat alle mengverhoudingen even waarschijnlijk zijn wanneer twee weefsels mengen in een voxel, op geen enkele basis rust.

Het EM-algoritme werd toegepast op gesimuleerde data en op 2-D sneden van reële MR-beelden van de hersenen. De resultaten op de gesimuleerde data toonden aan dat het histogram op zich vaak niet voldoende informatie bevat om de onderliggende modelparameters op een eenduidige manier te definiëren. Omdat het model zo complex is, kunnen er veel paramatersets bestaan die allemaal een goede benadering van het histogram opleveren. Wat de correcte oplossing van de andere onderscheidt, is dat ze resulteert in betekenisvolle classificaties van de beelden. Daarom is spatiale informatie tijdens de parameterschatting zelf vaak onontbeerlijk, en kan deze niet, zoals in een aantal bestaande methoden werd beschreven, simpelweg achteraf worden toegevoegd om de segmentaties te verbeteren. Experimenten op 2-D sneden van hoge-resolutie MR-beelden toonden aan dat het MRF-model de ruis in de segmentaties beduidend reduceert, en PV-voxels dwingt om op de rand tussen de samenstellende weefsels te liggen. Zonder zulk een MRF was de parameterschatting ondergedetermineerd voor beelden van lagere resolutie, maar de MRF-techniek kon niet worden getest op zulke beelden omdat onze huidige implementatie geen partieel volume in de richting loodrecht op het beeldvlak in rekening brengt. Verwacht wordt echter dat een stationair MRF kleine details in de segmentaties zou tenietdoen wanneer beelden van lagere resolutie worden bewerkt.

Er werd daarom besloten dat verbeterde 3-D spatiale modellen noodzakelijk zijn voor algemene, robuuste partieel-volumesegmentatie. Bijgevolg vormt het opstellen van realistische statistische modellen voor de vorm van de menselijke hersenen een belangrijke uitdaging voor verder onderzoek.

Preface

In such an interdisciplinary research field as medical imaging, writing a text that is meaningful and relevant to people with various scientific backgrounds, is not a trivial task. This thesis has resulted in the following papers:

Publication I: K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "Automated model-based bias field correction of MR images of the brain", *IEEE transactions on medical imaging*, vol. 18, no. 10, pp. 885-896, October 1999

Publication II: K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "Automated model-based tissue classification of MR images of the brain", *IEEE transactions on medical imaging*, vol. 18, no. 10, pp. 897-908, October 1999

Publication III: K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, P. Suetens: "Automated segmentation of multiple sclerosis lesions by model outlier detection", *IEEE transactions on medical imaging*, 2001 (accepted)

Publication IV: K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "A statistical framework for partial volume segmentation", *In preparation*

Rather than re-writing these and trying to mold them into one thick book, a summary is presented in which most of the mathematical specialities are left out in favor of a clearer presentation of the general ideas and the main contributions. The original papers, which form an integral part of this thesis, can be found after the summary. It is my hope that this somewhat unusual approach will make my work accessible to a broader public than just colleagues specializing in the same research topic.

1 Introduction

Several neuropathologies of the central nervous system such as multiple sclerosis (MS), schizophrenia, epilepsy, Alzheimer and Creutzfeldt-Jakob disease (CJD) are related to morphological and/or structural changes in the brain. Studying such diseases by objectively measuring these changes instead of assessing the clinical symptoms is of great social and economical importance. For variant CJD [1] for instance, quantifying changes in the brain would be useful to support diagnosis as there are no standard clinical diagnostic criteria for this disease to date [2]. In schizophrenia, measuring asymmetry of volume and shape of left versus right brain hemisphere structures could help to understand the pathogenesis of the disease [3]. In major clinical trials, aiming to evaluate whether a new therapy might favorably modify the evolution of a disease, directly assessing the reduction of clinical symptoms would be extremely expensive for such a slow and variable disease as MS [4]. Quantifying changes in the brain and correlating them to clinical findings could provide surrogate endpoints that may change quicker and be more reliable to measure. Such measurements could also offer important information for therapy planning and follow-up of individual patients.

Current medical imaging modalities, such as X-ray computed tomography, radionuclide emission tomography, ultrasound tomography and magnetic resonance imaging (MRI), make it possible to acquire quantitative information of human anatomy in three dimensions in a non-invasive way. Because of its high spatial resolution and its inherent high soft tissue contrast, MRI is superior to other imaging techniques for studying diseases of the nervous system. Traditionally, the interpretation of MR images is based on cross-sectional views printed on film, which only provides subjective qualitative or descriptive results. However, the digital character of MRI offers the opportunity to extract objective and accurate quantitative information from the images. Such quantitative measurements can be compared over time to assess morphological changes in one individual (for instance to monitor the progression of MS), between individuals, or between populations (for instance schizophrenics versus controls; drug versus placebo groups).

Manual analysis of the images by a trained human expert is a tedious and difficult task, because the structures of interest show complex edge structures in 3-D and may lack clearly visible anatomical borders. In clinical trials, the number of MR images is so large that manual analysis by human experts is extremely time-consuming. Furthermore, it is not clear how a human rater combines information obtained from the different channels when multi-spectral MR data are examined. Also, the intra- and inter-observer variability associated with manual delineations complicates the analysis of the results. This is exemplified by the interferon beta-1b trial in MS where there was a significant reduction in measured brain lesion load in the third year due to a systematic change in the manual tracings on MR images [5].

Therefore, there is a need for fully-automated methods for MR brain image quantification that can analyze large amounts of multi-spectral MR data in a reproducible way which correlates well with expert analyses. This thesis focuses on the automated quantification of pathology-related *intensity* abnormalities in MR brain

images of patients suffering from MS and CJD. While the developed techniques are general enough to be applicable to other neurological diseases with MRI signal changes as well, such as for instance leukomalacia, the analysis of *shape* abnormalities of brain substructures, as required for instance in epilepsy and schizophrenia, is out of the scope of this text.

1.1 Multiple sclerosis

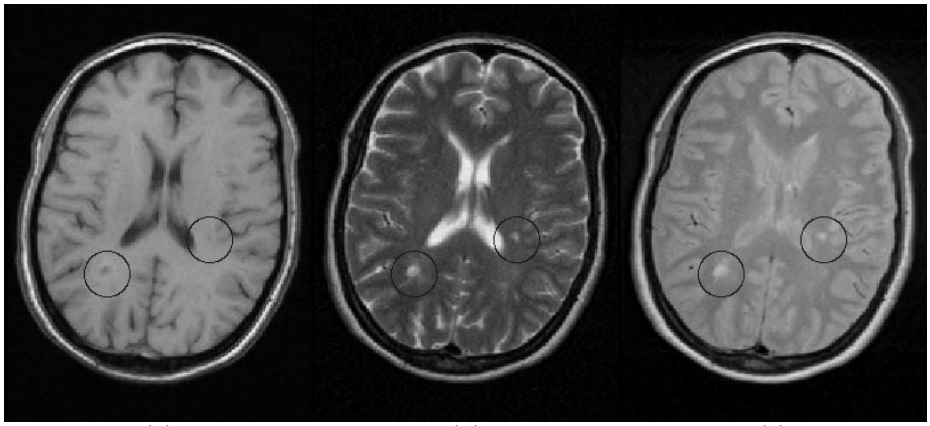
Multiple sclerosis is a common disease of young adults that primarily affects the white matter of the central nervous system. It typically begins in early adulthood, with peak age of onset at about 30 years [6], and has a variable prognosis. While 10 percent of patients do well for more than 20 years, 50 percent will need help walking within 15 years, and in rare cases, patients die within months [7]. The prevalence of MS varies considerably around the world, and is highest in northern Europe, southern Australia, and the middle part of North America. There are 250 000 to 350 000 patients with MS in the United States alone. While the cause of this disease is unknown, environmental factors and genetic susceptibility are believed to play a role [7].

Two major disease types are discriminated:

- relapsing-remitting MS is the type present in 80 percent of patients, and affects twice as many women as men. Typically, symptoms and signs evolve over a period of several days, stabilize, and then often improve. However, persistent signs of central nervous system disfunction may develop after a relapse, and the disease may progress between relapses. Recently, new therapies have been shown to reduce the short-term clinical manifestations, but data on long-term efficacy and safety are not available to date. Still, most patients continue to have relapses during treatment and ultimately become increasingly disabled [7].

- primary progressive MS affects 20 percent of patients, and is characterized by a gradually progressive clinical course and a similar incidence among men and women. There are no proven therapies for this type of MS [7].

MRI is known to be by far the best paraclinical test in MS, depicting abnormalities in 95 percent of patients [6]. Figure 1 shows a slice of a multi-spectral MR scan of a patient suffering from MS. MS lesions appear as regions with an increased signal intensity on the T2- and PD-weighted images, and an iso-intense or hypointense appearance in the T1-weighted channel. Today, diagnosis of MS is virtually always accompanied by a confirmatory MR study [6]. The risk of recurrence and the extent of disability can to some extent be predicted by the findings on MRI of the brain at the time of the first clinical episode [7]. Furthermore, MRI is increasingly being used to supplement traditional neurological disability scales such as the Extended Disability Status Scale (EDSS) [8] in assessing the progression of MS and in monitoring the effect of a new drug therapy. One of the reasons is that the EDSS is heavily weighted towards locomotor disability and that it has substantial intra-



(a) (b) (c)

Figure 1: Example MR brain scan of a patient suffering from MS: T1-weighted (a), T2-weighted (b), and PD-weighted image (c). MS lesions are encircled.

and inter-rater variability [9, 10]. MR measurements are far more sensitive and clearly reveal one important aspect of the underlying pathological process. They are therefore nowadays the primary outcome of preliminary clinical trials to evaluate whether a new therapy might favorably modify the evolution of the disease [9, 10]. A landmark study in this respect was the interferon beta-1b trial [5] that showed reduction in disease progression as assessed by MRI based findings.

The increasing use of MRI in diagnosing MS, in monitoring treatment, in studying the natural history of the disease and in clinical trials, poses the problem of obtaining accurate measurements of the lesions visible in the image data. The need for automated methods for MS lesion segmentation that can analyze large amounts of data in a reproducible way was recognized by the European Union, which established in 1996 the research project BIOMORPH [11]. The techniques developed in this thesis were applied to MR data from this project, which were acquired during a clinical trial in which fifty MS patients were repeatedly scanned with an interval of approximately one month over a period of about one year.

1.2 Creutzfeldt-Jakob disease

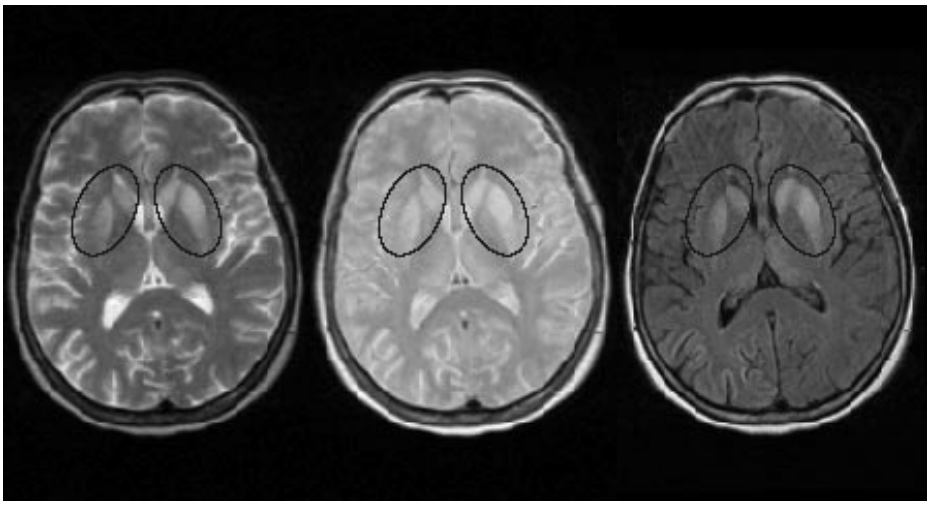
Creutzfeldt-Jakob disease is the major human form of the so-called transmissible spongiform encephalopathies, or prion diseases, a group of neurodegenerative diseases that affect human beings and animals. While cases of CJD are rare, it has received considerable attention because of its strange causative agent with novel modes of replication and transmission: an abnormal isoform of a normal protein, called prion protein (PrP). Furthermore, the appearance and rapid evolution to epidemic of bovine spongiform encephalopathy (BSE) in cattle ("mad cow disease") in the UK, its transmission to human beings, and its potentially severe threat to

public health, have received considerable media coverage.

There are different cases to be discriminated:

- 85 to 90 percent of cases of CJD are so-called sporadic CJD, a disease that primarily affects people between 50 and 70 years of age, and that causes death within 5 months on average. It occurs worldwide with an incidence of 0.5 to 1.5 cases per million population per year. Its cause remains a mystery [1].
- between 10 and 15 percent of CJD patients suffer from familial CJD, transmitted by inheritance of mutations in the PrP gene. In general, this form has an earlier age of onset and a more protracted course than sporadic CJD [1].
- iatrogenic CJD refers to CJD acquired by human-to-human transmission, tragically demonstrated by kuru, an epidemic in a remote mountainous area of New Guinea that was spread by ritual cannibalism. CJD has also been transmitted by contamination of routinely sterilized surgical instruments, by tissue donation, and by injection of human growth hormone [1].
- variant CJD is generally believed to be the human manifestation of BSE, a new disease in cattle that rose sharply from 16 cases in 1986 to 36 000 in the peak year 1992 in the UK. It is assumed that the epidemic started when the diets of calves and dairy cattle were supplemented with meat and bone meal produced by commercial rendering plants from scrapie-infected sheep or sporadic BSE cattle carcasses, and that it then spread further by the addition of material from infected cows to cattle feed [1, 12]. After 1992, the cases of BSE declined because of withdrawal of animal products from cattle feed and because of extensive slaughter of potentially infected animals. However, in 1995, 3 patients suffering from a new disease in humans, now called variant CJD, were reported in the UK, increasing to 85 confirmed cases at the date of writing. These patients were younger than those with the more familiar forms of the disease. Because of the long incubation periods seen in previous epidemics of human prion disease, many cases may follow in an epidemic that would be expected to evolve over decades [12].

To date, definitive diagnosis of CJD can still only be made by brain biopsy or necroscopy, or by tonsillar biopsy for variant CJD [2, 13]. While standard clinical diagnostic criteria allow reasonably accurate diagnosis in most of patients for sporadic, iatrogenic and familial CJD, no such criteria exist for variant CJD. Because subtle signal intensity abnormalities have been noted on MR brain scans in CJD patients, MRI has been suggested as a potential non-invasive test for early diagnosis of CJD. MR intensities in specific regions of the brain and intensity differences between such regions have been compared between CJD groups and groups of controls, by manually drawing regions of interest [14], or by subjective assessment of signal change on films [2]. Figure 2 shows an MR scan of a patient suffering from sporadic CJD; high signal can be seen in the deep gray matter structures. Accurate, automated quantification of such hyperintensities and correlating them with findings obtained from histology data is the aim of the EU project QAMRIC [15], from which the CJD data used in this thesis were obtained.



(a) (b) (c)

Figure 2: Example MR brain scan of a patient suffering from sporadic CJD: T2-weighted (a), PD-weighted (b), and FLAIR image (c). Signal abnormalities in the deep gray matter structures are encircled.

1.3 Segmentation methodology

Automated quantification of such signal intensity changes in MR images as shown in figures 1 and 2, is a very difficult problem. The complex 3-D shape of the brain and its affected areas, and ambiguities in the images induced by the imaging process, such as limited resolution, partial volume effects, noise, low contrast, intensity inhomogeneities and other artifacts, make interpretation difficult even for human experts. This is exemplified in MS, where there is a very wide variation in lesion volumes estimated by different experts, especially when these were trained at different institutions [16]. To complicate things further, the images used in this thesis are generally of a lower quality than what could be obtained nowadays in ideal circumstances with state-of-the art scanners. The MS scans are older longitudinal MRI data that were acquired during a clinical trial, with 5 mm slice thickness and a 0.5 mm gap between the slices. Although current MR technology allows for acquisition of high resolution image sequences, such valuable existing data cannot simply be disregarded. Because CJD in general and variant CJD in particular is such a rare disease, existing MR data originate from different centers with different scanners using different acquisition protocols, resulting in a wide variety of tissue contrast, voxel size and artifacts in the images. Furthermore, quality suffers from the need to keep the scanning time as short as possible for CJD patients to keep motion artifacts acceptable.

Low level segmentation methods, such as intensity thresholding, edge detection, region growing, region merging and morphological operations, are not well suited for automated quantification of the signal abnormalities addressed in this thesis.

These techniques rely entirely on image operators that analyze intensity, texture or shape locally in each voxel in the image, and are therefore too easily mislead by ambiguities in the images or require considerable user interaction. To overcome uncertainties in locally computed image features, models have been introduced by the medical imaging community that describe prior knowledge about the expected geometry and appearance of the anatomical objects of interest in the images:

- geometry-driven methods use the overall shape of an object to separate it from its surroundings in the image. Typically, a surface deforms under the influence of external image derived forces (attracting the surface to object edges etc.) and internal elasticity constraints (e.g. surface continuity and smoothness) [17]. An extensive survey of these methods in medical image analysis is given in [18]; recent examples include [19, 20, 21, 22]. Because of the wide variety in shape and position of MS lesions, such techniques would require considerable user interaction to provide a good initialization. Also, the notion of object boundary is not always meaningful in signal abnormalities in CJD, where intensities may change continuously within substructures of the brain (cf. figure 2).
- intensity-driven methods fit appropriate intensity models to the data, often explicitly taking imaging artifacts into account, such as the partial volume effect [23, 24, 25] and the intensity inhomogeneity present in MR images [26, 27, 28]. Voxels are classified independently or based on a Markov random field (MRF) prior [29]. Since the aim is to quantify changes in MR intensity, this is the methodology adopted in this thesis.

1.4 Mixture models and the EM algorithm

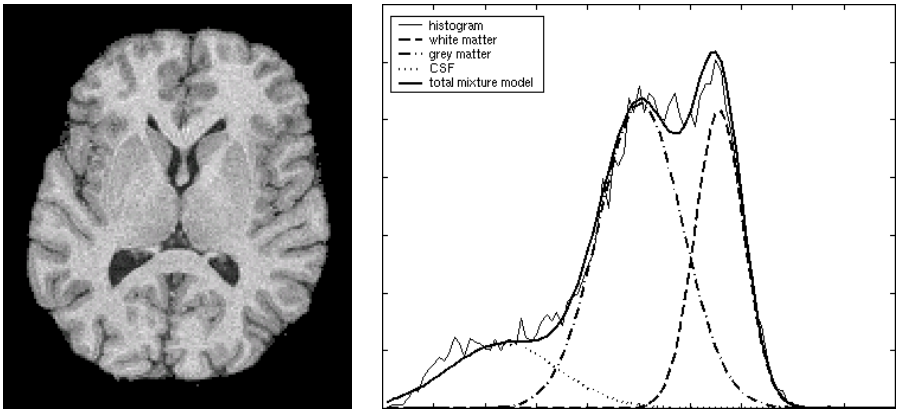
The intensity model that we use in this thesis is the so-called mixture of normal distributions [30, 31, 32]. Let $\boldsymbol{Y} = \{\boldsymbol{y}_j,\, j{=}1, 2, \ldots J\}$ be a C-channel MR image with a total of J voxels, where $\boldsymbol{y}_j$ denotes the possibly multi-spectral intensity of voxel j. Suppose that there are K tissue types present in the imaged area, and let $l_j{\in}\{1, 2, \ldots K\}$ denote the tissue type to which voxel j belongs. In the mixture model, it is assumed that each tissue type k has a typical intensity $\boldsymbol{\mu}_k$ in the image, with tissue-specific normally distributed intensity fluctuations in the voxels. In other words, the probability density that voxel j of tissue type l_j has intensity $\boldsymbol{y}_j$ is given by

$$f(\boldsymbol{y}_j \mid l_j, \Phi) = G_{\boldsymbol{\Sigma}_{l_j}}(\boldsymbol{y}_j - \boldsymbol{\mu}_{l_j}) \tag{1}$$

Here, $G_{\boldsymbol{\Sigma}}(\cdot)$ denotes a zero-mean normal distribution with covariance $\boldsymbol{\Sigma}$, and $\Phi = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k,\, k{=}1, 2, \ldots K\}$ represents the total set of model parameters.

For notational convenience, let all the voxel labels l_j be grouped in a label image $\boldsymbol{L} = \{l_j,\, j{=}1, 2, \ldots J\}$. It is assumed that the label l_j of each voxel is drawn independently from the labels of the other voxels, with an a priori known probability π_k, i.e.

$$f(\boldsymbol{L}) = \prod_j \pi_{l_j} \tag{2}$$



(a) (b)

Figure 3: The mixture model fitted to a T1-weighted brain MR image: the intracranial volume (a); and its intensity histogram with a mixture of normal distributions overlayed (b). The normal distributions correspond to white matter, gray matter, and CSF.

The overall probability density for image $\boldsymbol{Y}$ given the model parameters Φ is then given by

$$\begin{aligned} f(\boldsymbol{Y} \mid \Phi) &= \sum_{\boldsymbol{L}} f(\boldsymbol{Y} \mid \boldsymbol{L}, \Phi) f(\boldsymbol{L}) \\ &= \prod_{j} f(\boldsymbol{y}_j \mid \Phi) \\ \text{with} \quad f(\boldsymbol{y}_j \mid \Phi) &= \sum_{k} f(\boldsymbol{y}_j \mid l_j{=}k, \Phi) \cdot \pi_k \end{aligned} \tag{3}$$

Equation 3 is the well-known mixture model (see figure 3): it models the histogram of image intensities as a sum of normal distributions, each distribution weighted with its prior probability π_k.

Image segmentation aims to reconstruct the underlying tissue labels $\boldsymbol{L}$ based on the image $\boldsymbol{Y}$. If an estimation of the model parameters is somehow available, then each voxel can be assigned to the tissue type that best explains its intensity. Unfortunately, the result depends largely on the model parameters used. Typically, these are estimated by manually selecting representative points in the image of each of the classes considered. However, once all the voxels are classified, the model parameter estimation can in its turn automatically be improved based on all the voxels instead of on the subjectively selected ones alone. Intuitively, both the segmentation and the model parameters can be estimated in a more objective way by interleaving the segmentation with the estimation of the model parameters.

The Expectation-Maximization (EM) algorithm [33] formalizes this intuitive ap-

proach. It estimates the maximum likelihood (ML) parameters $\hat{\Phi}$

$$\hat{\Phi} = \arg\max_{\Phi} \log f(\boldsymbol{Y} \mid \Phi)$$

by iteratively filling in the unknown tissue labels $\boldsymbol{L}$ based on the current parameter estimation Φ, and recalculating Φ that maximizes the likelihood of the so-called complete data $\{\boldsymbol{Y}, \boldsymbol{L}\}$. More specifically, the algorithm interleaves two steps:

Expectation step : find the function

$$Q(\Phi \mid \Phi^{(m-1)}) = E_{\boldsymbol{L}}[\log f(\boldsymbol{Y}, \boldsymbol{L} \mid \Phi) \mid \boldsymbol{Y}, \Phi^{(m-1)}]$$

Maximization step: find

$$\Phi^{(m)} = \arg\max_{\Phi} Q(\Phi \mid \Phi^{(m-1)})$$

with m the iteration number. It has been shown that the likelihood $\log f(\boldsymbol{Y} \mid \Phi)$ is guaranteed to increase at each iteration for EM algorithms [34].

With the image model described above, the expectation step results in a statistical classification of the image voxels

$$f(l_j \mid \boldsymbol{Y}, \Phi^{(m-1)}) = \frac{f(\boldsymbol{y}_j \mid l_j, \Phi^{(m-1)}) \cdot \pi_{l_j}}{\sum_k f(\boldsymbol{y}_j \mid l_j = k, \Phi^{(m-1)}) \cdot \pi_k} \tag{4}$$

and the subsequent maximization step involves

$$\boldsymbol{\mu}_k^{(m)} = \frac{\sum_j f(l_j{=}k \mid \boldsymbol{Y}, \Phi^{(m-1)}) \cdot \boldsymbol{y}_j^{(m-1)}}{\sum_j f(l_j{=}k \mid \boldsymbol{Y}, \Phi^{(m-1)})} \tag{5}$$

$$\boldsymbol{\Sigma}_k^{(m)} = \frac{\sum_j f(l_j{=}k \mid \boldsymbol{Y}, \Phi^{(m-1)}) \cdot (\boldsymbol{y}_j^{(m-1)} - \boldsymbol{\mu}_k^{(m)}) \cdot (\boldsymbol{y}_j^{(m-1)} - \boldsymbol{\mu}_k^{(m)})^t}{\sum_j f(l_j{=}k \mid \boldsymbol{Y}, \Phi^{(m-1)})} \tag{6}$$

Thus, the algorithm iteratively improves the model parameters by interleaving two steps (see figure 4): classification of the voxels based on the estimation of the normal distributions (equation 4); and estimation of the normal distributions based on the classification (equations 5 and 6). Upon convergence, equation 4 yields the final classification result.

1.5 Overview and contributions

In this thesis the mixture model described in section 1.4 is extended in several consecutive steps in order to better represent real MR images of the brain. The overall strategy is to build increasingly better statistical models for normal brain MR images, and to detect MS- or CJD-related signal abnormalities as voxels that can not be well explained by the model. Figure 5 shows typical samples of the models constructed in this thesis in increasing order of complexity, starting from

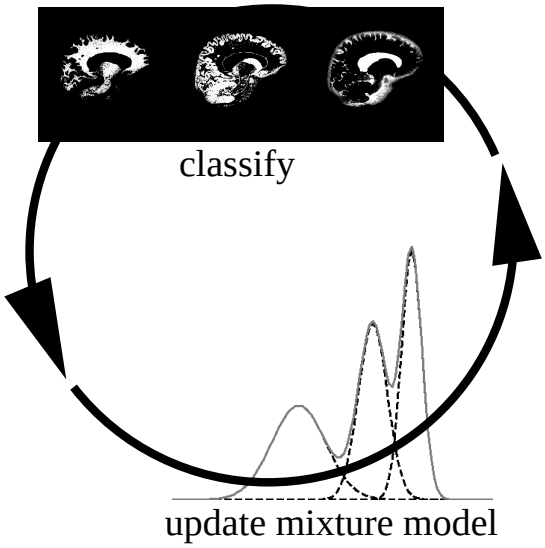


Figure 4: Estimating the model parameters of the mixture model with an EM algorithm results in an iterative 2-step process that interleaves classification of the voxels with re-estimation of the normal distributions.

the original mixture model, shown in figure 5 (a). For each of the models, the same EM framework is applied to estimate the model parameters. An unusual aspect of the work presented here is that all model parameters are estimated from the data itself, starting from an initialization that is obtained without user intervention. This ensures that the models retrain themselves fully automatically on each individual scan, allowing the method to analyze images with previously unseen MR pulse sequence and voxel size. In this way, subjective manual tuning and training is eliminated, which would make the results not fully reproducible and which would need to be re-done on almost each scan in the variable CJD data pool.

A first problem for the segmentation technique of section 1.4 is the corruption of MR images with a smoothly varying intensity inhomogeneity or bias field [35, 36], which results in a non-uniform intensity of the tissues over the image area as shown in figure 6. This bias is inherent to MR imaging and is caused by equipment limitations and patient-induced electrodynamic interactions [36]. Although not always visible for a human observer, it can cause serious misclassifications when intensity-based segmentation techniques are used. In chapter 2 the mixture model is therefore extended by explicitly including a parametric model for the bias field. Figure 5 (b) shows a typical sample of the resulting model. The model parameters are then iteratively estimated by interleaving three steps: classification of the voxels; estimation of the normal distributions; and estimation of the bias field. The algorithm is initialized with information from a digital brain atlas about the a priori expected location of tissue classes. This allows full automation of the method without need

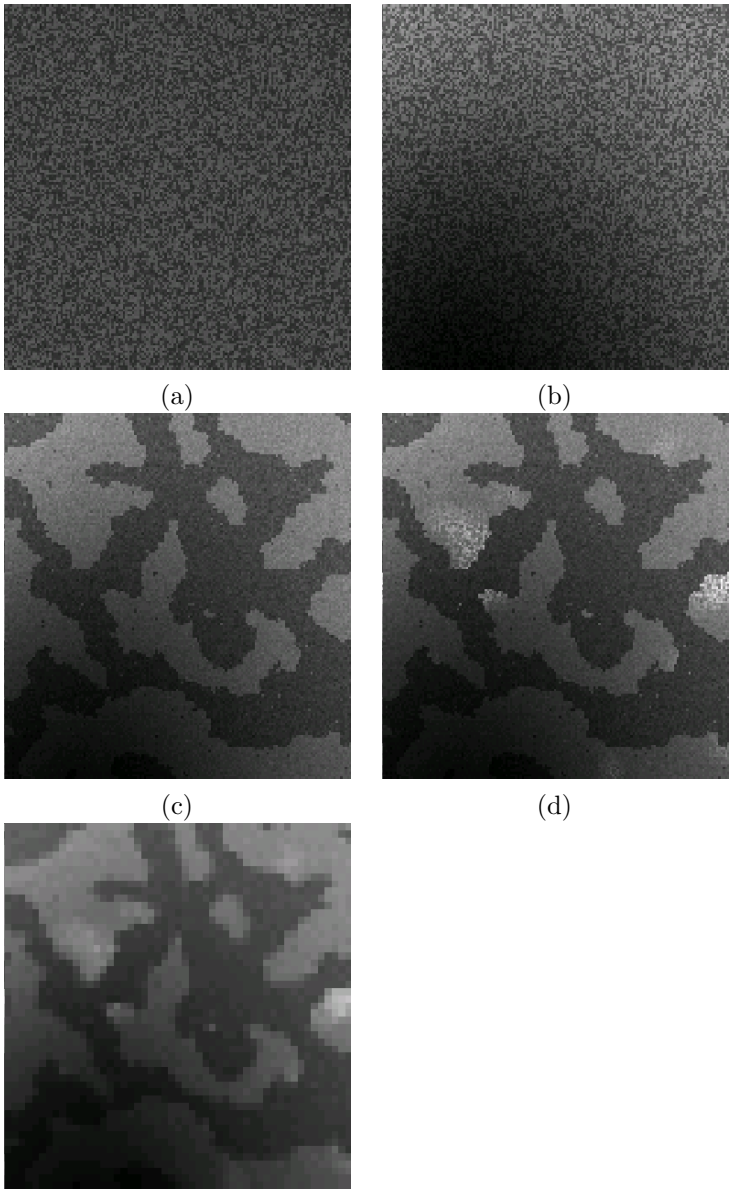


Figure 5: Illustration of the statistical models for brain MR images used in this thesis. The mixture model (a) is first extended with an explicit parametric model for the MR bias field (b). Subsequently, an improved spatial model is used that takes into account that tissues occur in clustered regions in the images (c). Then the presence of pathological tissues, which are not included in the statistical model, is considered (d). Finally, a downsampling step introduces partial voluming in the model (e).

for user interaction, yielding fully objective and reproducible results.

As a consequence of the assumption that the tissue type of each voxel is independent from the tissue type of the other voxels, each voxel is classified independently, only based on its intensity. However, the intensity of some tissues surrounding the brain is often very similar to that of brain tissue, which makes a correct classification based on intensity alone impossible. Therefore, the model is further extended in chapter 3 by introducing a MRF prior on the underlying tissue labels of the voxels. Such a MRF takes into account that the various tissues in brain MR images are not just randomly distributed over the image area, but occur in clustered regions of homogeneous tissue. This is illustrated in figure 5 (c), which shows a sample of the total resulting image model. The MRF brings general spatial and anatomical constraints into account during the classification, facilitating discrimination between tissue types with similar intensities such as brain and non-brain tissues.

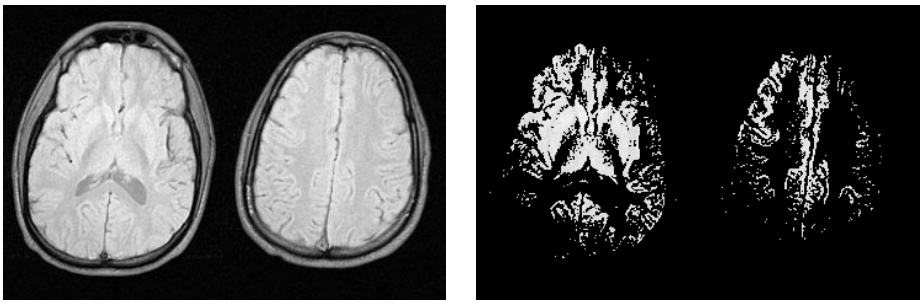
The method is further extended in chapter 4 in order to quantify MS lesions or CJD-related signal abnormalities in the images (see figure 5 (d)). Adding an explicit model for the pathological tissues is difficult because of the wide variety of their appearance in MR images, and because not every individual scan contains sufficient pathology for estimating the model parameters. These problems are circumvented by detecting lesions as voxels that are not well explained by the statistical model for normal brain MR images. Based on principles borrowed from the robust statistics literature, tissue-specific voxel weights are introduced that reflect the typicality of the voxels in each tissue type. Inclusion of these weights results in a robustized algorithm that simultaneously detects lesions as model outliers, and excludes these outliers from the model parameter estimation. In chapter 5, this outlier detection scheme is applied for fully automatic segmentation of MS lesions from brain MR scans. The method is validated by comparing the automatic lesion segmentations to manual tracings by human experts.

Chapter 6 investigates if the outlier detection technique can also be applied in CJD, where the MR intensity deviations are much more subtle than in MS. However, the model assumption that each voxel belongs to one single tissue type does not seem sufficiently accurate for this application. Due to the complex shape of the brain and the finite resolution of the images, a large part of the voxels lies on the border between two or more tissue types. Such border voxels are commonly referred to as partial volume (PV) voxels as they contain a mixture of several tissues at the same time. In order to be able to detect the subtle signal abnormalities in CJD, the model for normal brain MR images is further refined in chapter 7 by explicitly taking this PV effect into account. This is accomplished by introducing a downsampling step in the image model, adding up the contribution of a number of underlying subvoxels to form the intensity of a voxel. In voxels where not all subvoxels belong to the same tissue type, this causes partial voluming, as can be seen in figure 5 (e). The derived EM algorithm for estimating the model parameters provides a general framework for partial volume segmentation that encompasses and extends existing techniques. However, it is shown that the histogram alone often does not provide enough information to uniquely define the underlying parameters of such complex models. General robust partial volume segmentation therefore requires improved

models for the shape of the brain.

2 Automated bias field correction

The smoothly varying intensity inhomogeneity or bias field in MR images is not always visible to the human eye, as illustrated in figure 6. However, correcting the image intensities for bias field inhomogeneity is a necessary requirement for robust intensity analysis techniques. Early methods for bias field estimation and correction used phantoms to empirically measure the bias field inhomogeneity [37]. However, this approach assumes that the bias field is patient independent, which it is not [36]. Furthermore, it is required that the phantom's scan parameters are the same as the patient's, making this technique impractical and even useless as a retrospective bias correction method. In a similar vein, bias correction methods have been proposed for surface coil MR imaging using an analytic correction of the MR antenna reception profile [38], but these suffer from the same drawbacks as phantom based methods. Another approach, using homomorphic filtering [39], assumes that the frequency spectrum of the bias field and the image structures are well separated, but this assumption is generally not valid for MR images [37, 40].



(a) (b)

Figure 6: The MR bias field in a PD-weighted image. (a) shows two axial slices; (b) shows the same slices after intensity thresholding.

While bias field correction is needed for good segmentation, many approaches have exploited the idea that a good segmentation helps to estimate the bias field. Dawant *et al.* [40] manually selected some points inside white matter and estimated the bias field as the least-squares spline fit to the intensities of these points. They also presented a slightly different version where the reference points are obtained by an intermediate classification operation, using the estimated bias field for final classification. Meyer *et al.* [41] also estimated the bias field from an intermediate segmentation, but they allowed a region of the same tissue type to be broken up into several subregions which creates additional but sometimes undesired degrees of freedom.

Wells *et al.* [26] described an iterative method that interleaves classification with bias field correction based on ML parameter estimation using an EM algorithm. However, for each set of similar scans to be processed, their method, as well as its refinement by other authors [28, 27], needs to be supplied with specific tissue class

conditional intensity models. Such models are typically constructed by manually selecting representative points of each of the classes considered, which may result in segmentations that are not fully objective and reproducible.

In contrast, the method presented in Publication I does not require such a preceding training phase. Instead, a digital brain atlas is used with a priori probability maps for each tissue class to automatically construct intensity models for each individual scan being processed. This results in a fully automated algorithm that interleaves classification, bias field estimation and estimation of class-conditional intensity distribution parameters.

2.1 Image model and parameter estimation

The mixture model outlined in section 1.4 is extended to include a model for the bias field. We model the spatially smoothly varying bias fields as a linear combination of P polynomial basis functions $\phi_p(\boldsymbol{x}_j)$, p=$1, 2, \ldots P$, where $\boldsymbol{x}_j$ denotes the spatial position of voxel j. Not the observed intensities $\boldsymbol{y}_j$ but the bias corrected intensities $\boldsymbol{u}_j$ are now assumed to be distributed according to a mixture of class-specific normal distributions, such that equation 1 above is replaced by

$$\begin{aligned} f(\boldsymbol{y}_j \mid l_j, \Phi) &= G_{\boldsymbol{\Sigma}_{l_j}}(\boldsymbol{u}_j - \boldsymbol{\mu}_{l_j}) \\ \boldsymbol{u}_j &= \boldsymbol{y}_j - [\boldsymbol{b}_1 \ldots \boldsymbol{b}_C]^t \begin{bmatrix} \phi_1(\boldsymbol{x}_j) \\ \ldots \\ \phi_P(\boldsymbol{x}_j) \end{bmatrix} \end{aligned}$$

with $\boldsymbol{b}_c$, c=$1, 2, \ldots C$ indicating the bias field parameters of MR channel c, and $\Phi = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{b}_c,\ k$=$1, 2, \ldots K,\ c$=$1, 2, \ldots C\}$ as the total set of model parameters.

With the addition of the bias field model, estimation of the model parameters with an EM algorithm results in an iterative procedure that now interleaves 3 steps (see figure 7): classification of the image voxels (equation 4); estimation of the normal distributions (equations 5 and 6 but with the bias corrected intensities $\boldsymbol{u}_j$ replacing the original intensities $\boldsymbol{y}_j$); and estimation of the bias field. For the uni-spectral case, the bias field parameters are given by the following expression [1]:

$$\boldsymbol{b}^{(m)} = (\boldsymbol{A}^t \boldsymbol{W}^{(m)} \boldsymbol{A})^{-1} \boldsymbol{A}^t \boldsymbol{W}^{(m)} \boldsymbol{R}^{(m)}$$

with

$$\boldsymbol{A} = \begin{bmatrix} \phi_1(\boldsymbol{x}_1) & \phi_2(\boldsymbol{x}_1) & \phi_3(\boldsymbol{x}_1) & \ldots \\ \phi_1(\boldsymbol{x}_2) & \phi_2(\boldsymbol{x}_2) & \phi_3(\boldsymbol{x}_2) & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$\boldsymbol{W}^{(m)} = \mathrm{diag}(w_j^{(m)}), \quad w_j^{(m)} = \sum_k w_{jk}^{(m)}, \quad w_{jk}^{(m)} = \frac{f(l_j{=}k \mid \boldsymbol{Y}, \Phi^{(m-1)})}{\sigma_k^2}$$

[1] For more general expressions for the multi-spectral case we refer to Publication I; in the uni-spectral case $\boldsymbol{y}_j = y_j$, $\boldsymbol{\mu}_k = \mu_k$ and $\boldsymbol{\Sigma}_k = \sigma_k^2$

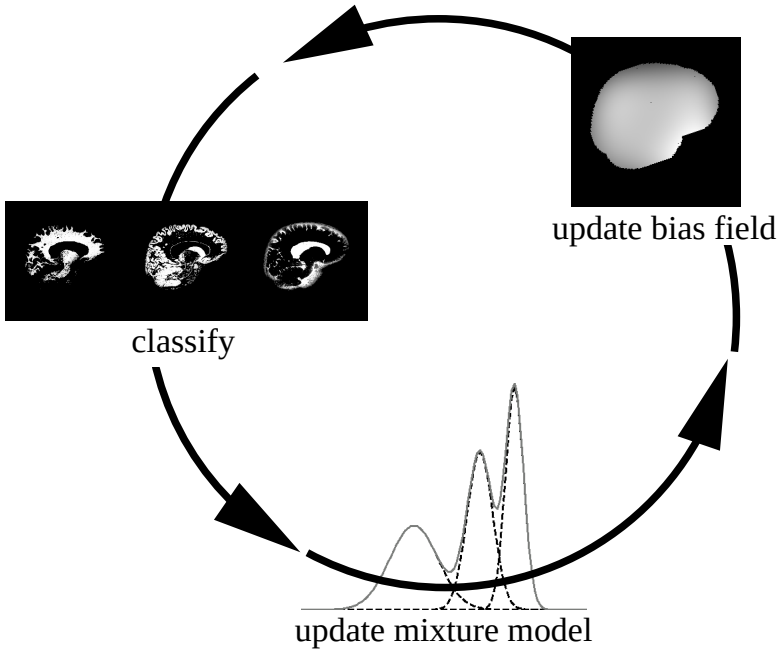


Figure 7: Adding a model for the bias field results in a 3-step EM algorithm that iteratively interleaves classification, estimation of the normal distributions, and bias field correction.

$$\boldsymbol{R}^{(m)} = \begin{bmatrix} y_1 - \tilde{y}_1^{(m)} \\ y_2 - \tilde{y}_2^{(m)} \\ \vdots \end{bmatrix}, \quad \tilde{y}_j^{(m)} = \frac{\sum_k w_{jk}^{(m)} \mu_k^{(m)}}{\sum_k w_{jk}^{(m)}}$$

This can be interpreted as follows (see figure 8). Based on the current classification and distribution estimation, a prediction $\{\tilde{y}_j, j{=}1, 2, \ldots J\}$ of the MR image without the bias field is constructed (figure 8 (b)). A residue image $\boldsymbol{R}$ (figure 8 (c)) is obtained by subtracting this predicted signal from the original image (figure 8 (a)). The bias (figure 8 (e)) is then estimated as a weighted least-squares fit through the residue image using the weights $\boldsymbol{W}$ (figure 8 (d)), each voxel's weight being inversely proportional to the variance of the class that voxel belongs to. As can be seen from figure 8 (d), the bias field is therefore computed primarily from voxels that belong to classes with a narrow intensity distribution, such as white and gray matter. The smooth spatial model extrapolates the bias field from these regions, where it can be confidently estimated from the data, to regions where such estimate is ill-conditioned (CSF, non-brain tissues).

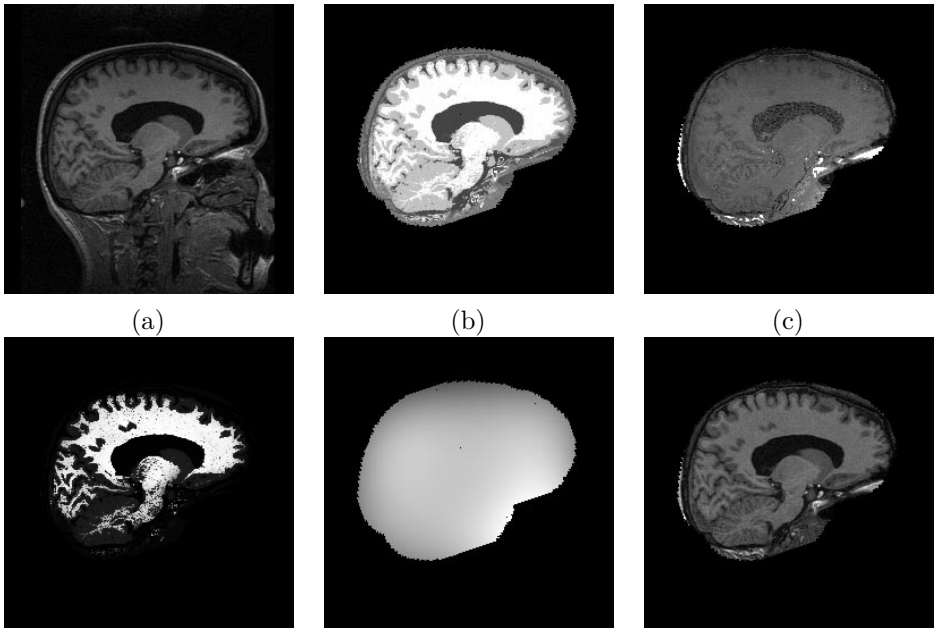
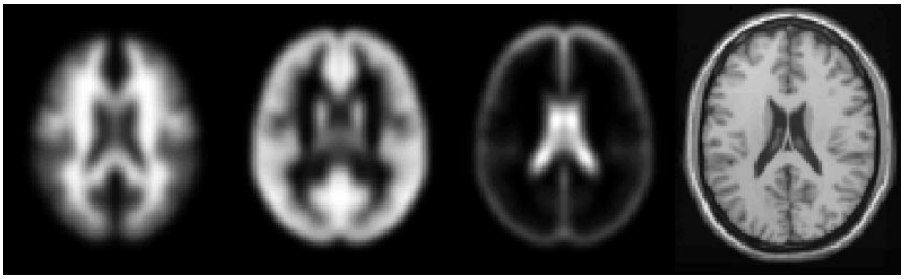


Figure 8: Illustration of the bias correction step on a 2D slice of a T1-weighted MR image: (a) original image; (b) predicted signal based on previous iterations; (c) residue image; (d) weights; (e) estimated bias field; and (f) corrected image. (Source: Publication I)

2.2 Initialization using a digital brain atlas

The iterative model parameter estimation algorithm described above needs to be initialized with a first estimate of the parameters. One possibility to obtain such an estimate, is to have the user manually select voxels that are representative for each of the classes considered. However, to eliminate the variability induced by such a preceding training phase, we avoid manual intervention by the use of a digital brain atlas that contains spatially varying prior probability maps for the location of white matter, gray matter, and CSF (see figure 9). These probability maps were obtained by averaging binary white matter, gray matter and CSF segmentations of MR brain images from a large number of subjects, after normalization of all images into the same space using an affine transformation [42]. To apply this a priori information, the atlas is first normalized to the space of the study image by matching the study image to a T1 template associated with the atlas (see figure 9) using the affine multi-modality registration technique based on maximization of mutual information of Maes *et al.* [43]. It has been shown [43] that this registration criterion is fairly insensitive to moderate bias fields, such that it can be applied fully automatically and reliably to the uncorrected MR images. The properly registered and reformatted a priori tissue probability maps of the atlas provide an initial esti-



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

Figure 9: Digital brain atlas with spatially varying a priori probability maps for (a) white matter, (b) gray matter and (c) CSF. High intensities indicate high a priori probabilities. The atlas also contains a T1 template image (d) which is used for registration of the study images to the space of the atlas. (Source: Publication I)

mate of the classification from which initial values for the class-specific distribution parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ can be computed. This approach frees us from having to interactively indicate representative voxels of each class, which makes our method more objective and reproducible and allows the method to be fully automated.

The classification, bias field parameters and intensity distribution parameters are then updated using the iterative scheme based on the EM procedure as outlined above. During iterations, the atlas is further used to spatially constrain the classification by assigning its prior probability maps to the a priori class probabilities π_k. Thus, the voxels are not only classified based on their intensities, but also based on their spatial position. This makes the algorithm more robust, especially when the images are corrupted with a heavy bias field.

2.3 Examples and discussion

Examples of the performance of the method are shown in figures 10 and 11. Figure 10 depicts the classification of a high-resolution sagittal T1-weighted MR image, both for the original 2-step algorithm without bias correction of section 1.4, and for the new 3-step algorithm with bias correction. Because a relatively strong bias field reduces the intensities at the top of the head, bias correction is necessary as white matter is wrongly classified as gray matter in that area otherwise. Figure 11 clearly shows the efficiency of the bias correction on a 2-D multi-slice T1-weighted image. Such multi-slice images are acquired in an interleaved way, and are typically corrupted with a slice-by-slice constant intensity offset, commonly attributed to gradient eddy currents and crosstalk between slices [35], and clearly visible as an interleaved bright-dark intensity pattern in a cross-section orthogonal to the slices in figure 11.

In earlier EM approaches for bias correction, the class-specific intensity distribution parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ were determined by manual training and kept fixed during the iterations [26, 28, 27]. It has been reported [28, 44, 45] that these meth-

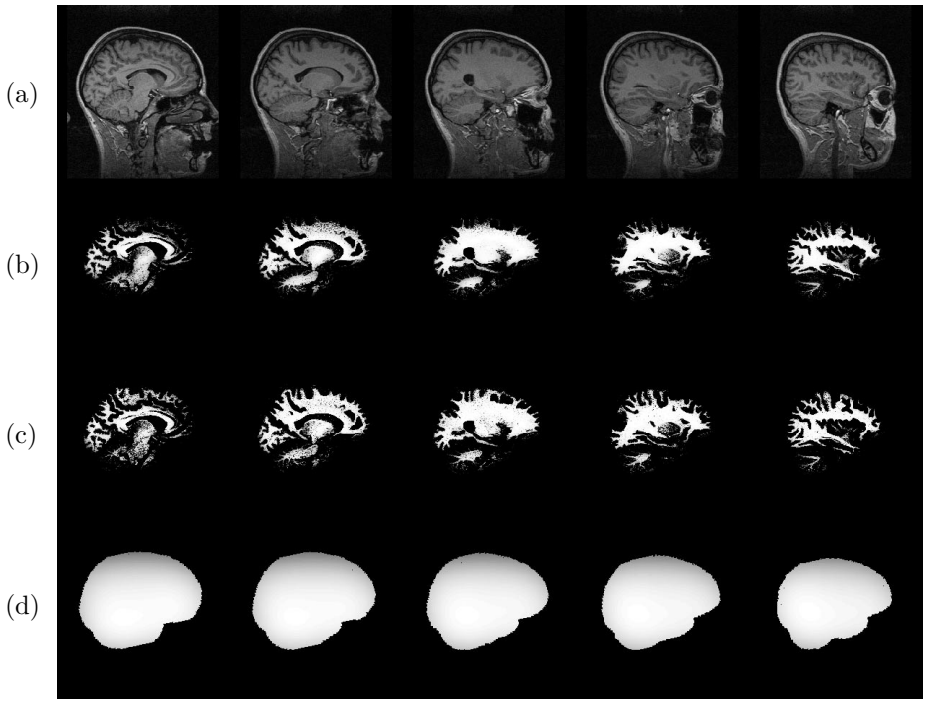


Figure 10: Slices of a high-resolution T1-weighted MR image illustrating the performance of the method: (a) original data; (b) white matter classification without bias correction; (c) white matter classification with bias correction; (d) estimated bias field. (Source: Publication I)

ods are sensitive to the training of the classifier, i.e. they produce different results depending on which voxels were selected for training. In contrast, our algorithm estimates its tissue-specific intensity distributions fully automatically on each individual scan being processed, starting from a digital brain atlas. This avoids all manual intervention, yielding fully objective and reproducible results. Moreover, it eliminates the danger of overlooking some tissue types during a manual training phase, which is typically a problem in regions surrounding the brain, consisting of several different tissues, and which may cause severe errors in the residue image and the bias field estimation [28, 45]. Guillemaud and Brady [28] proposed to model non-brain tissues by a single class with a uniform distribution, artificially assigning the non-brain tissue voxels a zero weight for the bias estimation. This is not necessary with our algorithm: the class distribution parameters are updated at each iteration from all voxels in the image and classes consisting of different tissue types are automatically assigned a large variance. Since the voxel weights for the bias correction are inversely proportional to the variance of the class each voxel is classified to, such tissues are therefore automatically assigned a low weight for the bias estimation.

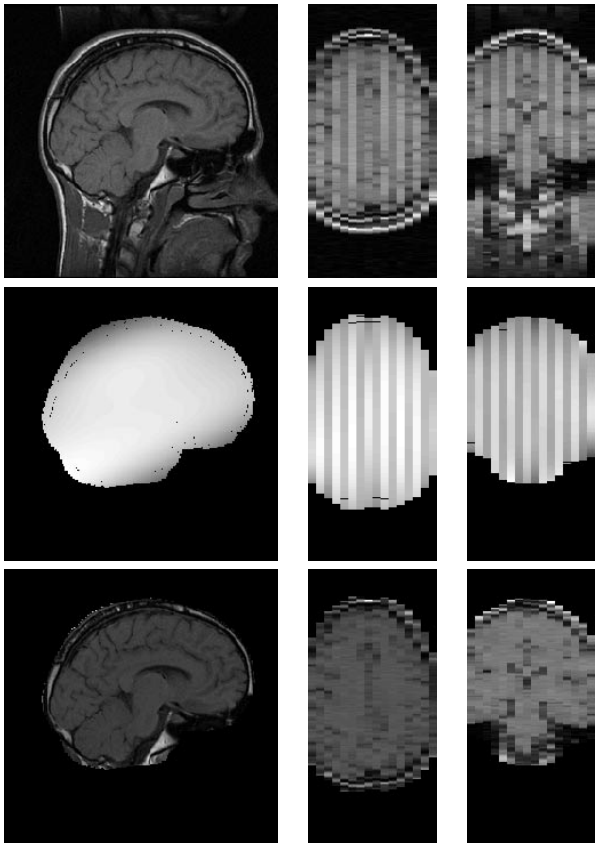


Figure 11: An example of bias correction of a T1-weighted 2-D multi-slice image corrupted with slice-by-slice offsets. From top row to bottom row: original data, estimated bias, corrected data. (Source: Publication I)

A number of authors have proposed bias correction methods that do not use an intermediate classification. Styner, Brechbühler *et al.* [44, 46] find the bias field for which as many voxels as possible have an intensity in the corrected image that lies close to that of a number of predefined tissue types. Other approaches search for the bias field that makes the histogram of the corrected image as sharp as possible [47, 48, 49]. The method of Sled *et al.* [47] for instance is based on deconvolution of the histogram of the measured signal, assuming that the histogram of the bias field is Gaussian, while Mangin [48] and Likar *et al.* [49] use entropy to measure the histogram sharpness. Contrary to our approach, these methods treat all voxels alike for bias estimation. This looks rather unnatural, since it is obvious that the white matter voxels, which have a narrow intensity histogram, are much more suited for bias estimation than for instance the tissues surrounding the brain or ventricular CSF. As argued above, our algorithm takes this explicitly into account by the class-dependent weights assigned to each voxel. Furthermore, lesions can

be so large in a scan of a MS patient that the histogram of the corrected image may be sharpest when the estimated bias field follows the anatomical distribution of the lesions. As will be shown in chapter 4, our method can be made robust for the presence of such pathologic tissues in the images, estimating the bias field only from normal brain tissue.

3 Modeling spatial context

As a consequence of the assumption that the tissue type of each voxel is independent from the tissue type of the other voxels, each voxel is classified independently based on its intensity. This yields acceptable classification results as long as the different classes are well separated in intensity feature space, i.e. have a clearly discernible associated intensity distribution. Unfortunately, this is not always true for MR images of the brain, especially not when only one MR channel is available. Whereas white matter, gray matter and CSF usually have a characteristic intensity, voxels surrounding the brain often show an MR intensity that is very similar to brain tissue. This may result in erroneous classifications of small regions surrounding the brain as gray matter or white matter.

In Publication II, the model was therefore further extended with a MRF model, introducing general spatial and anatomical constraints during the classification. The MRF is designed to facilitate discrimination between brain and non-brain tissues while preserving the detailed interfaces between the various tissue classes within the brain.

3.1 Regularization using a MRF model

Previously, it was assumed that the label l_j of each voxel is drawn independently from the labels of the other voxels, leading to equation 2 for the prior probability distribution for the underlying label image $\boldsymbol{L}$. Now a more complex model for $\boldsymbol{L}$ is used, more specifically a Markov random field. Such a MRF model assumes that the probability that voxel j belongs to tissue type k depends on the tissue type of its neighbors. The Hammersley-Clifford theorem states that such a random field is a Gibbs random field (see [29] and the references therein), i.e. its configurations obey a Gibbs distribution

$$f(\boldsymbol{L} \mid \Phi) = Z(\Phi)^{-1} \exp[-U(\boldsymbol{L} \mid \Phi)]$$

where $Z(\Phi) = \sum_{\boldsymbol{L}} \exp[-U(\boldsymbol{L} \mid \Phi)]$ is a normalization constant called the partition function, and $U(\boldsymbol{L} \mid \Phi)$ is an energy function dependent on the model parameters Φ.

A simple MRF is used that is defined on a so-called first order neighborhood system, i.e. only the six nearest neighbors on the 3-D-image lattice are used. Let $\mathcal{N}_j^p$ denote the set of the four neighbors of voxel j in the plane and $\mathcal{N}_j^o$ its two neighbors out of the plane. Since the voxel size in the z-direction is usually different from the within-plane voxel size in MR images, the following Potts model (the extension of the binary Ising model [50] to more than 2 classes) is used:

$$U(\boldsymbol{L} \mid \Phi) = \frac{1}{2} \sum_j \left(\sum_{j' \in \mathcal{N}_j^p} \xi_{l_j l_{j'}} + \sum_{j' \in \mathcal{N}_j^o} \nu_{l_j l_{j'}} \right)$$

Here, the MRF parameters $\xi_{kk'}$ and $\nu_{kk'}$ denote the cost associated with transition from class k to class k' among neighboring voxels in the plane and out of the plane,

respectively. If these costs are higher for neighbors belonging to different classes than for neighbors of the same tissue class, the MRF favors configurations of $\boldsymbol{L}$ where each tissue is spatially clustered. An example of this was shown in figure 5 (c).

It has been described in the literature that fine structures, such as the interface between white matter and gray matter in brain MR images, can be erased by the Potts/Ising MRF model [27, 51]. The MRF may over-regularize such subtle borders and tempt to produce nicely smooth interfaces. Therefore, a modification is used that penalizes anatomically impossible combinations such as a gray matter voxel surrounded by non-brain tissues, while at the same time preserving edges between tissues that are known to border each other. We impose that a voxel surrounded by white matter and gray matter voxels must have the same a priori probability for white matter as for gray matter by adding appropriate constraints on the MRF transition costs $\xi_{kk'}$ and $\nu_{kk'}$. As a result, voxels surrounded by brain tissues have a low probability for CSF and other non-brain tissues, and a high but equal probability for white and gray matter. The actual decision between white and gray matter is therefore only based on the intensity, so that the interface between white and gray matter is unaffected by the MRF. Similar constraints are applied for other interfaces as well.

Since the voxel labels are not independent with this model, the expectation step of the EM algorithm no longer yields the classification of equation 4. Due to the interaction between the voxels, the exact calculation of $f(l_j \mid \boldsymbol{Y}, \Phi^{(m-1)})$ involves calculation of all the possible realizations of the MRF, which is not computationally feasible. Therefore, an approximation was adopted that was proposed by Zhang [52] and Langan *et al.* [53], based on the mean field theory from statistical mechanics. This mean field approach suggests an approximation to $f(l_j \mid \boldsymbol{Y}, \Phi^{(m-1)})$ based on the assumption that the influence of $l_{j'}$ of all other voxels $j' \neq j$ in the calculation of $f(l_j \mid \boldsymbol{Y}, \Phi^{(m-1)})$ can be approximated by the influence of their classification $f(l_{j'} \mid \boldsymbol{Y}, \Phi^{(m-2)})$ from the previous iteration. This yields

$$f(l_j \mid \boldsymbol{Y}, \Phi^{(m-1)}) \approx \frac{f(\boldsymbol{y}_j \mid l_j, \Phi^{(m-1)}) \cdot \pi_j^{(m-1)}(l_j)}{\sum_k f(\boldsymbol{y}_j \mid l_j = k, \Phi^{(m-1)}) \cdot \pi_j^{(m-1)}(k)} \tag{7}$$

for the classification, where

$$\pi_j^{(m-1)}(l_j) \sim \exp\Bigg(- \sum_{j' \in \mathcal{N}_j^p} \sum_k f(l_{j'}{=}k \mid \boldsymbol{Y}, \Phi^{(m-2)}) \cdot \xi_{l_j k} - \\ - \sum_{j' \in \mathcal{N}_j^o} \sum_k f(l_{j'}{=}k \mid \boldsymbol{Y}, \Phi^{(m-2)}) \cdot \nu_{l_j k} \Bigg)$$

The difference with equation 4 lies herein, that now the a priori probability that a voxel belongs to a specific tissue class depends on the classification of its neighboring voxels.

With the addition of the MRF, the subsequent maximization step in the EM algorithm not only involves updating the intensity distributions and recalculating

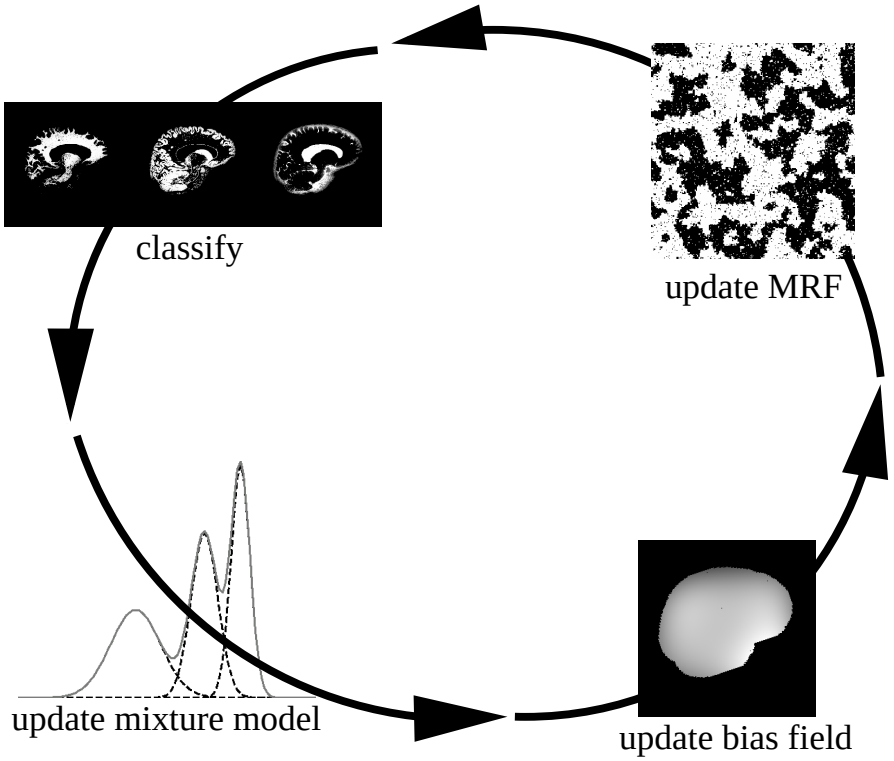
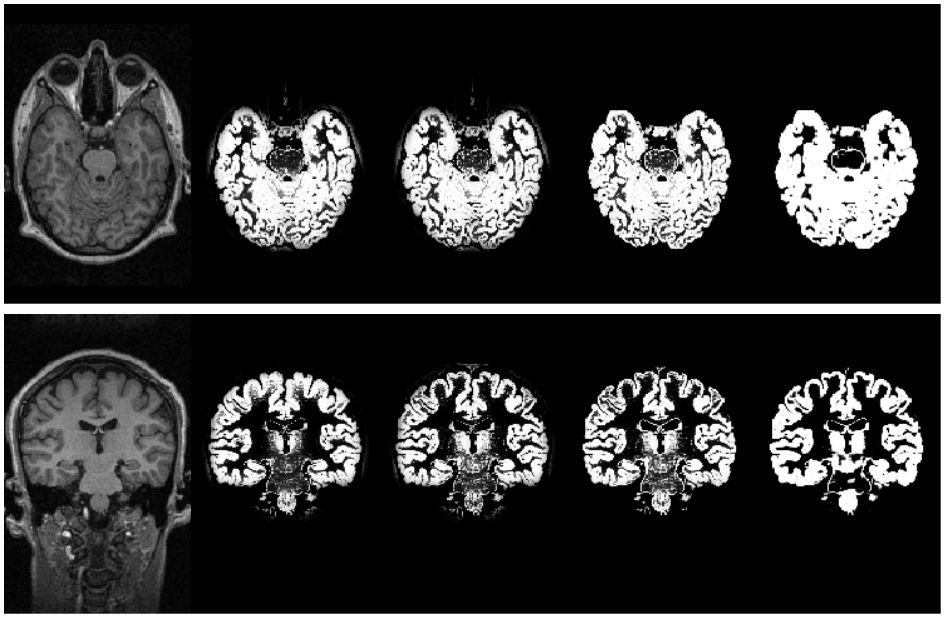


Figure 12: The extension of the model with a MRF prior results in a 4-step algorithm that interleaves classification, estimation of the normal distributions, bias field correction, and estimation of the MRF parameters.

the bias field, but also estimating the MRF parameters $\{\xi_{kk'}\}$ and $\{\nu_{kk'}\}$. As a result, the total iterative scheme now consists in 4 steps, shown in figure 12. The calculation of the MRF parameters poses a difficult problem for which a heuristic, non-iterative approach is used. For each neighborhood configuration $(\mathcal{N}^p, \mathcal{N}^o)$, the number of times that the central voxel belongs to class k in the current classification is compared to the number of times it belongs to class k', for every couple of classes (k, k'). This results in an overdetermined linear system of equations that is solved for the MRF parameters $(\xi_{kk'}, \nu_{kk'})$ using a least squares fit procedure [29].

3.2 Example

Figure 13 demonstrates the influence of each component of the algorithm on the resulting segmentations of a T1-weighted image. First, the method of chapter 2, where each voxel is classified independently, was used without the bias correction step (figure 13 (b)). It can be seen that white matter at the top of the brain



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

Figure 13: Example of how the different components of the algorithm work. From left to right: T1-weighted image (a), gray matter segmentation without bias field correction and MRF (b), gray matter segmentation with bias field correction but without MRF (c), gray matter segmentation with bias field correction and MRF (d), gray matter segmentation with bias field correction and MRF without constraints (e). (Source: Publication II)

is misclassified as gray matter. This was clearly improved when the bias field correction step was added (figure 13 (c)). However, some tissues surrounding the brain have intensities that are similar to brain tissue, and are wrongly classified as gray matter. With the MRF model described in section 3.1, a better distinction is obtained between brain tissues and tissues surrounding the brain (figure 13 (d)). This is most beneficial in case of single-channel MR data, where it is often difficult to differentiate such tissues only based on their intensity. The MRF cleans up the segmentations of brain tissues, while preserving the detailed interface between gray and white matter, and between gray matter and CSF. Figure 14 depicts a 3-D volume rendering of the gray matter segmentation map when the MRF is used.

To demonstrate the effect of the constraints on the MRF parameters $\xi_{kk'}$ and $\nu_{kk'}$ described in section 3.1, the same image was processed without such constraints (figure 13 (e)). The resulting segmentation shows nicely distinct regions, but small details, such as small ridges of white matter, are lost. The MRF prior has over-regularized the segmentation and should therefore not be used in this form.

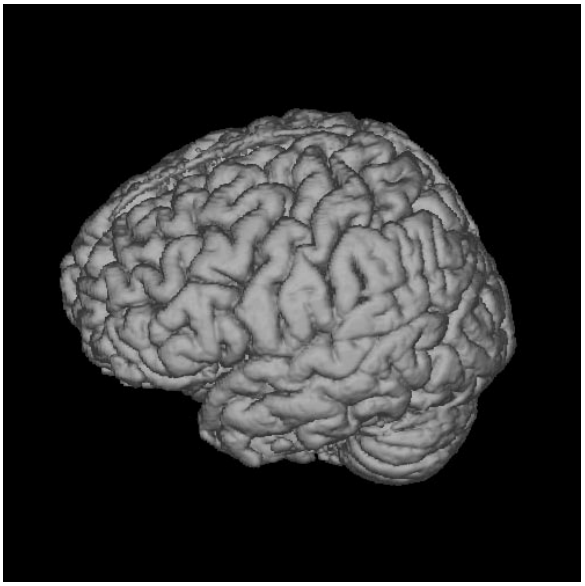


Figure 14: 3-D volume rendering of the gray matter segmentation of the data of figure 13 with bias field correction and MRF. (Source: Publication II)

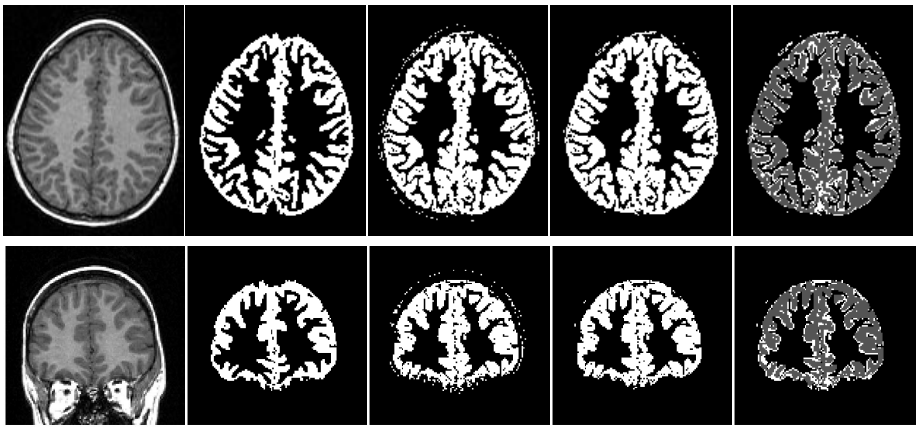
3.3 Validation and conclusions

The method was validated on simulated MR images of the head that were generated by the BrainWeb MR simulator [54], for varying number of MR channels, noise and severity of the bias fields. The automatic segmentations of each tissue k were compared with the known ground truth by calculating the similarity index

$$\frac{2V_{12}^k}{V_1^k + V_2^k} \tag{8}$$

where V_{12}^k denotes the volume of the voxels classified as tissue k by both raters, and V_1^k and V_2^k the volume of class k assessed by each of the raters separately. This metric, first described by Dice [55] and recently re-introduced by Zijdenbos *et al.* [56], attains the value of 1 if both segmentations are in full agreement, and 0 if there is no overlap at all. For all the simulated data, it was found that the total brain volume was accurately segmented, but the segmentation of gray matter and white matter individually did generally not attain the same accuracy. This was caused by misclassification of the white matter/gray matter interface, where PV voxels do not belong to either white matter or gray matter, but are really a mixture of both.

The automated method was also validated by comparing its segmentations of nine brain MR scans of children to the manual tracings by a human expert. The automated and manual segmentations showed an excellent similarity index of 95% on average for the total brain, but a more moderate similarity index of 83% for



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

Figure 15: Comparison between manual delineation and automated tissue classification on a representative dataset. From left to right: axial and coronal slice (a), corresponding manual segmentation of gray matter (b), automatic segmentation of gray matter without MRF prior (c), automatic segmentation of gray matter with MRF (d), difference between manual and automatic segmentation with MRF shown in white (e). (Source: Publication II)

gray matter. Figure 15 depicts the location of misclassified gray matter voxels for a representative dataset. It can be seen that the automatic algorithm segments the gray matter-CSF interface in more detail than the manual tracer. Some tissue surrounding the brain is still misclassified as gray matter, although this error is already reduced compared to the situation where no MRF prior is used. However, by far most misclassifications are due to the classification of gray-white matter PV voxels to gray matter by the automated method. The human observer has segmented white matter consistently as a thicker structure than the automatic algorithm.

Partial voluming violates the model assumption that each voxel belongs to only one single class. In reality, PV voxels are a mixture of tissues and every segmentation method that tries to assign them exclusively to one class is doomed to fail. The problem is especially important in images of the brain since the interface between gray and white matter is highly complex, which results in a high volume of PV voxels compared to the volume of pure tissue voxels. Misclassification of this thin interface gives therefore immediate rise to considerable segmentation errors [57]. In chapter 7, the model will therefore be extended to explicitly include partial voluming.

4 Model outliers and robust parameter estimation

So far, only the segmentation of MR images of normal brains has been addressed. In order to quantify MS lesions or CJD-related signal abnormalities in the images, the method needs to be further extended. Adding an explicit model for the pathological tissues is difficult because of the wide variety of their appearance in MR images, and because not every individual scan contains sufficient pathology for estimating the model parameters. In Publication III, these problems are circumvented by detecting lesions as voxels that are not well explained by the statistical model for normal brain MR images. The method simultaneously detects the lesions as model outliers, and excludes these outliers from the model parameter estimation.

4.1 Background

Suppose that J samples $\boldsymbol{y}_j$, j=1, 2, ... J are drawn independently from a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ that are grouped in $\Phi = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ for notational convenience. Given these samples, the ML parameters Φ can be assessed by maximizing

$$\sum_j \log f(\boldsymbol{y}_j \mid \Phi) \tag{9}$$

which yields

$$\begin{aligned} \boldsymbol{\mu} &= \frac{\sum_j \boldsymbol{y}_j}{J} \\ \boldsymbol{\Sigma} &= \frac{\sum_j (\boldsymbol{y}_j - \boldsymbol{\mu})(\boldsymbol{y}_j - \boldsymbol{\mu})^t}{J} \end{aligned} \tag{10}$$

In most practical applications, however, the assumed normal model is only an approximation to reality, and estimation of the model parameters Φ should not be severely affected by the presence of a limited amount of model outliers. Considerable research efforts in the field of robust statistics [58] have resulted in a variety of methods for robust estimation of model parameters in the presence of outliers, from which the so-called M-estimators [58] present the most popular family.

Considering equation 9, it can be seen that the contribution to the log-likelihood of an observation that is atypical of the normal distribution is high, since $\lim_{f(\boldsymbol{y}|\Phi)\to 0} \log f(\boldsymbol{y} \mid \Phi) = -\infty$. The idea behind M-estimators is to alter equation 9 slightly in order to reduce the effect of outliers. A simple way to do this, which has recently become very popular in image processing [59] and medical image processing [31, 28, 32, 60], is to model a small fraction of the data as being drawn from a rejection class that is assumed to be uniformly distributed. It can be shown that assessing the ML parameters is now equivalent to maximizing

$$\sum_j \log\big(f(\boldsymbol{y}_j \mid \Phi) + \lambda\big), \qquad \lambda \geq 0 \tag{11}$$

with respect to the parameters Φ, where λ is an a priori chosen threshold [59]. Since $\lim_{f(\boldsymbol{y}|\Phi)\to 0} \log(f(\boldsymbol{y} \mid \Phi) + \lambda) = \log(\lambda)$, the contribution of atypical observations on the log-likelihood is reduced compared to equation 9.

One possibility to numerically maximize equation 11, is to calculate iteratively the weights

$$t(\boldsymbol{y}_j \mid \Phi^{(m-1)}) = \frac{f(\boldsymbol{y}_j \mid \Phi^{(m-1)})}{f(\boldsymbol{y}_j \mid \Phi^{(m-1)}) + \lambda} \tag{12}$$

based on the parameter estimation $\Phi^{(m-1)}$ in iteration $(m-1)$, and subsequently update the parameters $\Phi^{(m)}$ accordingly:

$$\begin{aligned} \boldsymbol{\mu}^{(m)} &= \frac{\sum_j t(\boldsymbol{y}_j \mid \Phi^{(m-1)}) \cdot \boldsymbol{y}_j}{\sum_j t(\boldsymbol{y}_j \mid \Phi^{(m-1)})} \\ \boldsymbol{\Sigma}^{(m)} &= \frac{\sum_j t(\boldsymbol{y}_j \mid \Phi^{(m-1)}) \cdot (\boldsymbol{y}_j - \boldsymbol{\mu}^{(m)})(\boldsymbol{y}_j - \boldsymbol{\mu}^{(m)})^t}{\sum_j t(\boldsymbol{y}_j \mid \Phi^{(m-1)})} \end{aligned} \tag{13}$$

Solving an M-estimator by iteratively re-calculating weights and updating the model parameters based on these weights, is commonly referred to as the W-estimator [61]. The weight $t(\boldsymbol{y}_j \mid \Phi) \in [0, 1]$ reflects the typicality of sample i with respect to the normal distribution. For typical samples, $t(\boldsymbol{y}_j \mid \Phi) \simeq 1$, whereas $t(\boldsymbol{y}_j \mid \Phi) \simeq 0$ for samples that deviate far from the model. Comparing equation 13 with equation 10, it can therefore be seen that the M-estimator effectively down-weights observations that are atypical for the normal distribution, making the parameter estimation more robust against such outliers.

4.2 From typicality weights to outlier belief values

Since each voxel j has only a contribution of $t(\boldsymbol{y}_j \mid \Phi)$ to the parameter estimation, the remaining fraction

$$1 - t(\boldsymbol{y}_j \mid \Phi) \tag{14}$$

reflects the belief that it is a model outlier. The ultimate goal in our application is to identify these outliers as they are likely to indicate pathological tissues. However, the dependence of equation 14 through $t(\boldsymbol{y}_j \mid \Phi)$ on the determinant of the covariance matrix $\boldsymbol{\Sigma}$ prevents its direct interpretation as a true outlier belief value.

In statistics, an observation $\boldsymbol{y}$ is said to be abnormal with respect to a given normal distribution if its so-called Mahalanobis-distance $d = \sqrt{(\boldsymbol{y} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})}$ exceeds a predefined threshold. Regarding equation 12, the Mahalanobis distance at which the belief that a voxel is an outlier exceeds the belief that it is a regular sample, decreases with increasing $|\boldsymbol{\Sigma}|$. Therefore, the Mahalanobis distance threshold above which voxels are considered abnormal, changes over the iterations as $\boldsymbol{\Sigma}$ is updated. Because of this problem, it is not clear how λ should be chosen.

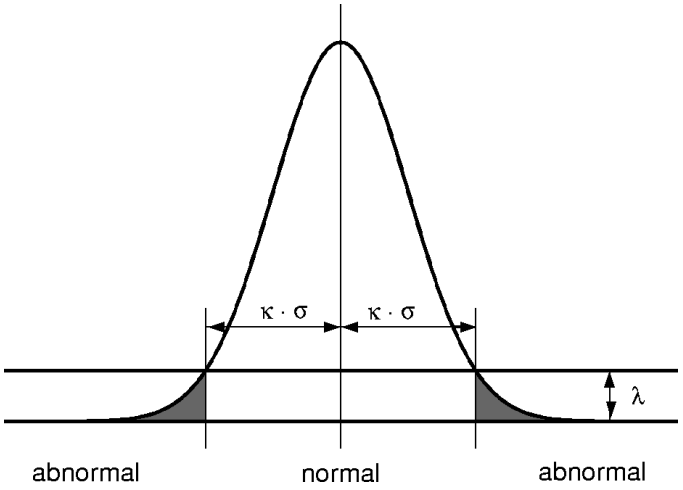


Figure 16: The threshold κ defines the Mahalanobis distance at which the belief that a voxel is a model outlier exceeds the belief that it is a regular sample (this figure depicts the uni-spectral case, where $\mathbf{\Sigma} = \sigma^2$).

Therefore, equation 12 is modified into

$$t(\boldsymbol{y}_j \mid \Phi^{(m-1)}) = \frac{f(\boldsymbol{y}_j \mid \Phi^{(m-1)})}{f(\boldsymbol{y}_j \mid \Phi^{(m-1)}) + \frac{1}{\sqrt{(2\pi)^C |\mathbf{\Sigma}^{(m-1)}|}} \exp(-\frac{1}{2}\kappa^2)}$$

where $|\mathbf{\Sigma}|$ is explicitly taken into account, and where λ is re-parameterized using the more easily interpretable κ. This $\kappa \geq 0$ is an explicit Mahalanobis-distance threshold that specifies a statistical significance level, as illustrated in figure 16. The lower κ is chosen, the easier voxels are considered as outliers. On the other hand, choosing $\kappa = \infty$ results in $t(\boldsymbol{y}_j \mid \Phi^{(m-1)}) = 1, \forall j$ which causes no outliers to be detected at all.

4.3 Robust estimation of MR model parameters

Based on the same concepts, the EM framework used in the previous chapters for estimating the parameters of models for normal brain MR images can be extended to detect model outliers such as MS lesions or CJD-related signal abnormalities. In the original EM algorithm, a statistical classification $f(l_j \mid \boldsymbol{Y}, \Phi^{(m-1)})$ is performed in the expectation step, and the subsequent maximization step involves updating the model parameters according to this classification. The weights $f(l_j{=}k \mid \boldsymbol{Y}, \Phi^{(m-1)})$, $k{=}1, 2, \ldots K$ represent the degree to which voxel j belongs to each of the K tissues. However, since $\sum_k f(l_j{=}k \mid \boldsymbol{Y}, \Phi^{(m-1)}) = 1$, an observation that is atypical for each of the normal distributions cannot have a small membership value for all tissue types simultaneously.

A similar approach as the one described above, where equation 9 was replaced with the more robust equation 11 and solved with a W-estimator, results in a maximization step in which model outliers are down-weighted. The resulting equations for updating the model parameters are identical to the original ones, provided that the weights $f(l_j \mid \boldsymbol{Y}, \Phi^{(m-1)})$ are replaced everywhere with a combination of two weights $f(l_j \mid \boldsymbol{Y}, \Phi^{(m-1)}) \cdot t(\boldsymbol{y}_j \mid l_j, \Phi^{(m-1)})$, where

$$t(\boldsymbol{y}_j \mid l_j, \Phi^{(m-1)}) = \frac{f(\boldsymbol{y}_j \mid l_j, \Phi^{(m-1)})}{f(\boldsymbol{y}_j \mid l_j, \Phi^{(m-1)}) + \frac{1}{\sqrt{(2\pi)^C |\boldsymbol{\Sigma}_k^{(m-1)}|}} \exp(-\frac{1}{2}\kappa^2)} \tag{15}$$

reflects the degree of typicality of voxel j in tissue class l_j. Since $\sum_k f(l_j = k \mid \boldsymbol{Y}, \Phi^{(m-1)}) \cdot t(\boldsymbol{y}_j \mid l_j = k, \Phi^{(m-1)})$ is not constrained to be unity, model outliers can have a small degree of membership in all tissue classes simultaneously. Therefore, observations that are atypical for each of the K tissue types, have a reduced weight on the parameter estimation, which robustizes the EM-procedure. Upon convergence of the algorithm, the belief that voxel j is a model outlier is given by

$$1 - \sum_k f(l_j{=}k \mid \boldsymbol{Y}, \Phi) \cdot t(\boldsymbol{y}_j \mid l_j{=}k, \Phi) \tag{16}$$

Chapter 5 discusses the use of this outlier detection scheme for fully automated segmentation of MS lesions from brain MR images, while the application of the method to quantification of CJD-related signal abnormalities is addressed in chapter 6.

5 Application to multiple sclerosis

In Publication III, the outlier detection scheme of chapter 4 was applied for fully automatic segmentation of MS lesions from brain MR scans that consist of T1-, T2- and PD-weighted images. Unfortunately, outlier voxels also occur outside MS lesions. This is typically true for partial volume voxels that, in contravention to the assumptions made, do not belong to one single tissue type but are rather a mixture of more than one tissue. Since they are perfectly normal brain tissue, though, they are prevented from being detected as MS lesion by introducing constraints on intensity and context on the weights $t(\boldsymbol{y}_j \mid l_j, \Phi)$ calculated in equation 15.

5.1 Intensity and contextual constraints

- Since MS lesions appear hyper-intense on both the PD- and the T2-weighted images, only voxels that are brighter than the mean intensity of gray matter in these channels are allowed to be outliers.

- Since around 90–95 % of the MS lesions are white matter lesions, the contextual constraint is added that MS lesions should be located in the vicinity of white matter. In each iteration, the normal white matter is fused with the lesions to form a mask of the total white matter. Using a MRF as in chapter 3, a voxel is discouraged from being classified as MS lesion in the absence of neighboring white matter. Since the MRF parameters are estimated from the data in each iteration as in chapter 3, these contextual constraints automatically adapt to the voxel size of the data.

The complete method is summarized in figure 17. It iteratively interleaves statistical classification of the voxels into normal tissue types, assessment of the belief for each voxel that it is not part of an MS lesion based on its intensity and on the classification of its neighboring voxels, and, only based on what is considered as normal tissue, estimation of the MRF, intensity distributions and bias field parameters. Upon convergence, the belief that voxel j is part of an MS lesion is obtained by equation 16. The method is fully automated, with only one single parameter that needs to be experimentally tuned: the Mahalanobis threshold κ in equation 15.

5.2 Validation

As part of the BIOMORPH project [11], we analyzed MR data acquired during a clinical trial in which fifty MS patients were repeatedly scanned with an interval of approximately one month over a period of about one year. The serial image data consisted at each time point of a PD/T2-weighted image pair and a T1-weighted image with 5 mm slice thickness. From 10 of the patients, 2 consecutive time points were manually analyzed by a human expert who traced MS lesions based only on the T2-weighted images. The automatic algorithm was repeatedly applied with values of the Mahalanobis distance κ varying from 2.7 (corresponding to a significance level

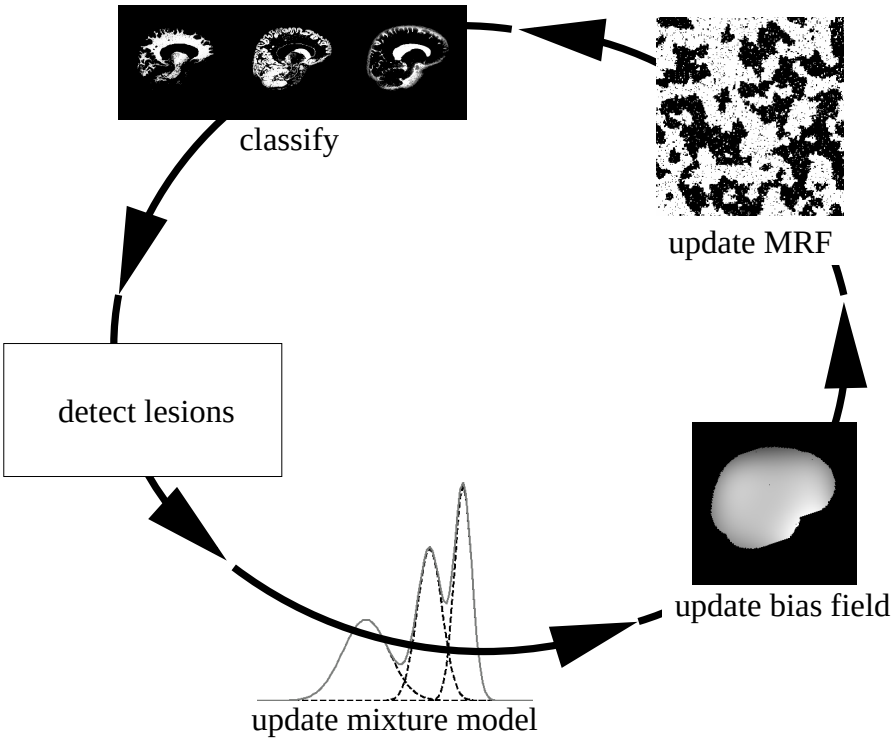
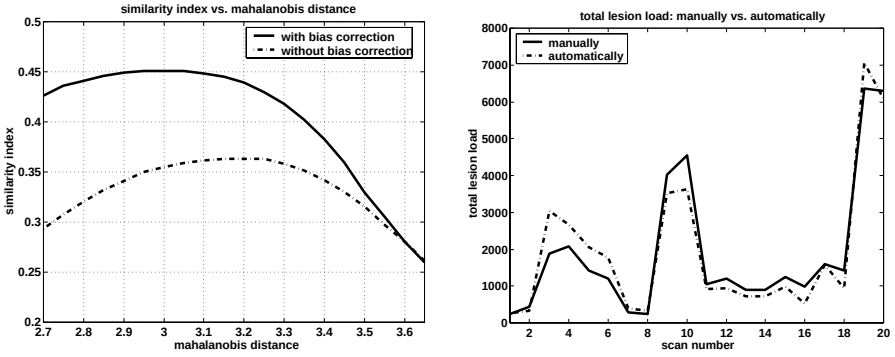


Figure 17: The complete method for MS lesion segmentation iteratively interleaves classification of the voxels into normal tissue types, MS lesion detection, estimation of the normal distributions, bias field correction and MRF parameter estimation.

of p=0.063) to 3.65 (corresponding to p=0.004), in steps of 0.05. The automatic delineations were compared with the expert segmentations by comparing the so-called total lesion load (TLL), measured as the number of voxels that were classified as MS lesion, on these 20 scans. The TLL value calculated by the automated method decreased when κ was increased, since the higher κ, the less easily voxels are rejected from the model. Varying κ from 2.7 to 3.65 resulted in an automatic TLL of respectively 150% to only 25% of the expert TLL. However, despite the strong influence of κ on the absolute value of the TLL, the linear correlation between the automated TLL's of the 20 scans and the expert TLL's was remarkable insensitive to the choice of κ. Over this wide range, the correlation coefficient varied between 0.96 and 0.98.

Comparing the TLL of two raters does not take into account any spatial correspondence of the segmented lesions. We therefore calculated the similarity index defined in equation 8, which is simply the volume of intersection of the two segmentations divided by the mean of the two segmentation volumes. For the 20 scans, figure 18 (a) depicts the value of this index for varying κ's, both with and without



(a) (b)

Figure 18: Similarity index between the automatic and the expert lesion delineations on 20 images for varying κ's, with and without the bias field correction component enabled in the automated method (a); the 20 automatic total lesion load measurements for $\kappa = 3$ shown along with the expert measurements (b). (Source: Publication III)

the bias correction step included in the algorithm, clearly demonstrating the need for bias field correction. The best correspondence, with a similarity index of 0.45, was found for $\kappa \simeq 3$. For this value of κ, the automatic TLL was virtually equal to the expert TLL, and therefore, a similarity index of 0.45 means that less than half of the voxels labeled as lesion by the expert were also identified by the automated method, and vice versa.

For illustration purposes, the expert TLL's of the 20 scans are depicted along with the automatic ones for $\kappa = 3$ in figure 18 (b). A paired t-test did not reveal a significant difference between the manual and these automatic TLL ratings (p=0.94). Scans 1 and 2 are two consecutive scans from one patient, 3 and 4 from the next and so on. Note that in 9 out of 10 cases, the two ratings agree over the direction of the change of the TLL over time. Figure 19 displays the MR data of what is called scan 19 in figure 18 (b) and the automatically calculated classification along with the lesion delineations performed by the human expert.

5.3 Discussion

Most of the methods for MS lesion segmentation described in the literature are semi-automated rather than fully-automated methods, designed to facilitate the tedious task of manually outlining lesions by human experts, and to reduce the inter- and intra-rater variability associated with such expert segmentations. Typical examples of user interaction in these approaches include accepting or rejecting automatically computed lesions [62], or manually drawing regions of pure tissue types for training an automated classifier [63, 64, 65, 62]. While these methods have proven to be useful, they remain impractical when hundreds of scans need to be analyzed as part

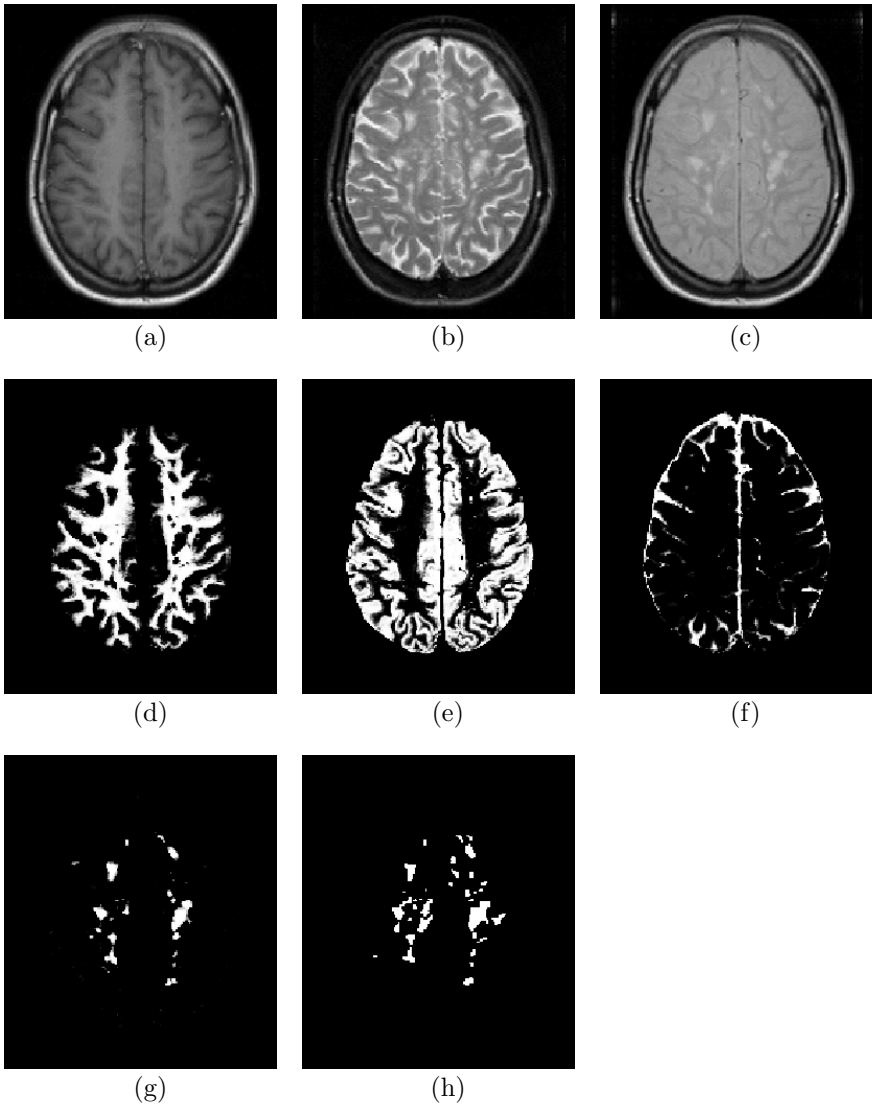


Figure 19: Automatic classification of one of the 20 serial data sets that were also analyzed by a human expert. T1-weighted image (a); T2-weighted image (b); PD-weighted image (c); white matter classification (d); gray matter classification (e); CSF classification (f); MS lesion classification (g); and expert delineation of the MS lesions (h). (Source: Publication III)

of a clinical trial, and the variability of manual tracings is not totally removed. In contrast, the method presented here is fully automated as it uses a probabilistic brain atlas to train its classifier. Furthermore, the atlas provides spatial information that avoids non-brain voxels from being classified as MS lesion, making the method work without the often-used tracing of the intracranial cavity in a preprocessing step [63, 64, 65, 62, 66, 67].

A unique feature of our algorithm is that it automatically adapts its intensity models and contextual constraints when analyzing images that were acquired with a different MR pulse sequence or voxel size. Zijdenbos *et al.* described [68] and validated [16] a fully automated pipeline for MS lesion segmentation based on an artificial neural network classifier. Similarly, Kikinis, Guttmann *et al.* [66, 69] have developed a method with minimal user intervention that is built on the EM classifier of Wells *et al.* [26] with dedicated pre- and post-processing steps. Both methods use a fixed classifier that is only trained once and that is subsequently used to analyze hundreds of scans. In clinical trials, however, inter-scan variations in cluster shape and location in intensity space cannot be excluded, not only because of hardware fluctuations of MR scanners over a period of time, but also because different imagers may be used in a multi-center trial [69]. In contrast to the methods described above, our algorithm retrains its classifier on each individual scan, making it adaptive to such contrast variations.

Often, a post-processing step is applied to automatically segmented MS lesions, in which false positives are removed based on a set of experimentally tuned morphologic operators, connectivity rules, size thresholds, etc [63, 64, 66]. Since such rules largely depend on the voxel size, they may need to be re-tuned for images with a different voxel size. Alternatively, images can be re-sampled to a specific image grid before processing, but this introduces partial voluming that can reduce the detection of lesions considerably, especially for small lesion loads [69]. To avoid these problems, we have added explicit contextual constraints on the iterative MS lesions detection that automatically adapt to the voxel size. Similar to other methods [68, 63, 65, 67], we exploit the knowledge that the majority of MS lesions occurs inside white matter. Our method fuses the normal white matter with the lesions in each iteration, producing, in combination with MRF constraints, a prior probability mask for white matter that is automatically updated during the iterations. Since the MRF parameters are re-estimated for each individual scan, the contextual constraints automatically adapt to the voxel size of the images.

Although the algorithm we present is fully automatic, an appropriate Mahalanobis distance threshold κ has to be chosen in advance. When evaluating the role of κ, a distinction has to be made between the possible application areas of the method. In clinical trials, the main requirement for an automated method is that its measurements change in response to a treatment in a manner proportionate to manual measurements, rather than having an exact equivalence in the measurements [9, 10]. In section 5.2, it was shown that the automatic measurements always kept changing proportionately to the manual measurements for a wide range of κ, with high correlation coefficients between 0.96 and 0.98. Therefore, the actual choice of κ is fairly unimportant for this type of application. However, the role of

κ is much more critical when the goal is to investigate the basic MS mechanisms or time correlations of lesion groups in MS time series, as these applications require that the lesions are also spatially correctly detected. In general, the higher the resolution and the better the contrast between lesions and unaffected tissue in the images, the easier MS lesions are detected by the automatic algorithm and the higher κ should be chosen. Therefore, the algorithm presumably needs to be tuned for different studies, despite the automatic adaptation of the tissue models and the MRF parameters to the data.

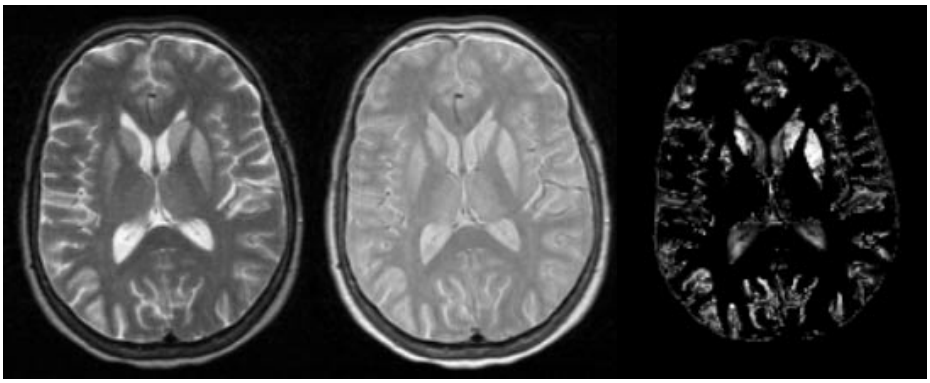
6 Application to Creutzfeldt-Jakob disease

As part of the QAMRIC project [15], it was tested if the outlier detection scheme of chapter 4 could also be applied for automated quantification of the subtle intensity deviations that are often observed in MR scans of patients suffering from CJD. In contrast to the MS data sets described in chapter 5, that were all acquired in a standardized way using the same scanner with the same acquisition parameters, the collected CJD MR data originate from different centers due to the rareness of the disease. Since these centers use different scanners with different acquisition protocols, the QAMRIC data pool exhibits a wide variety of available MR channels, tissue contrast, voxel size and artifacts in the images. Because the scanning time needs to be kept as short as possible for CJD patients in order to keep motion artifacts acceptable, the resolution of the scans is low, with a slice thickness of 6 mm in most cases. Although it has been suggested that diffusion-weighted images and fluid attenuated inversion recovery (FLAIR) images are more sensitive in showing CJD related signal abnormalities than PD/T2-weighted images [2, 70, 71], these images were only available for a limited number of cases. Therefore, the automated method was applied on PD/T2-weighted image pairs only since these were acquired for most patients.

Figure 20 shows an example of a PD/T2-weighted scan of a patient suffering from sporadic CJD, along with the outlier belief values that were automatically calculated with the technique described in chapter 4. Figure 21 shows the same for an image of a variant CJD case, acquired at a different imaging center. The slice thickness is 3 mm in the first case, and 5 mm in the second. It can be seen that the algorithm detects abnormal signals in the deep gray matter structures in both cases. However, considerable amounts of partial volume voxels also have high outlier belief values, predominantly along the interface between gray matter and CSF. For the segmentation of MS lesions, such false positives were suppressed by a MRF that discourages MS lesions from being detected outside white matter as explained in section 5.1. While this type of "cleaning up" of classifications works satisfactorily well for the problem of MS lesion segmentation, we have not been successful in discriminating PV voxels from truly affected voxels in CJD, where the intensity deviations are much more subtle. Moreover, in most CJD images, no signal abnormalities were detected at all.

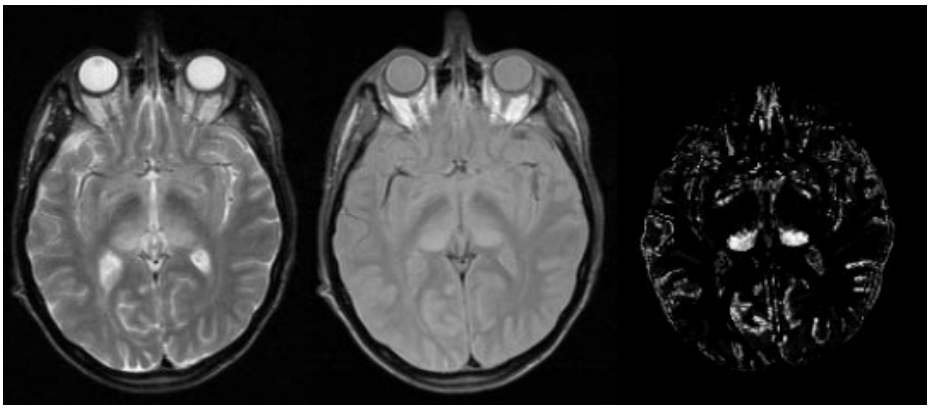
Figure 22 depicts the scatter plot of the intracranial volume of the scan of figure 20 after bias correction. The shape of the estimated normal distributions for white matter, gray matter and CSF is also indicated by drawing the ellipse corresponding to a Mahalanobis distance of 2.5. The smallest ellipse corresponds to white matter, and the largest to CSF. Clearly, the partial voluming arising from the large slice thickness of the images, results in dramatically large covariances for gray matter and especially for CSF. In fact, a very large part of the intensities inside the intracranial volume can be well explained by the normal distribution for CSF, and therefore it is very difficult to discriminate subtle signal abnormalities from normal brain tissue.

It is tempting to try to model the PV effect explicitly. This would avoid PV



(a) (b) (c)

Figure 20: Outlier detection on a brain MR scan of a patient suffering from sporadic CJD. T2-weighted image (a); PD-weighted image (b); outlier belief values (c).



(a) (b) (c)

Figure 21: Outlier detection on a brain MR scan of a patient suffering from variant CJD. T2-weighted image (a); PD-weighted image (b); outlier belief values (c).

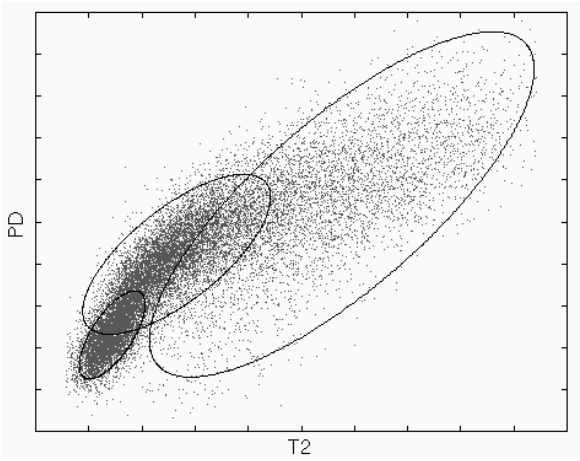


Figure 22: Scatter plot of the intracranial volume of the scan shown in figure 20, illustrating the difficulties caused by the PV effect. The ellipses indicate a Mahalanobis distance of 2.5 in the estimated normal distributions of white matter (smallest ellipse), gray matter (intermediate ellipse) and CSF (largest ellipse). The dramatically large size of the estimated covariances, especially the one for CSF, makes detection of subtle signal abnormalities very difficult.

voxels from being detected as model outliers on the one hand, and it would provide a more realistic model that fits more tightly to the data, making it easier to detect subtle intensity deviations on the other hand. In chapter 7, the model will therefore be further extended to include partial voluming.

7 Partial volume segmentation

As shown in chapters 3 and 6, the assumption that the voxels in MR images belong to only one single tissue type causes considerable classification errors. In Publication IV, the image model was therefore extended with a downsampling step, introducing partial voluming. The resulting EM algorithm provides a general framework for partial volume segmentation that encompasses and extends existing techniques.

7.1 Downsampling and resulting image model

To recapitulate, the model used so far assumed that a label image $\boldsymbol{L}$, indicating the tissue type l_j of each voxel j, j=1, 2, ... J, is sampled from a probability distribution $f(\boldsymbol{L} \mid \Phi)$. Subsequently, tissue-dependent normally distributed noise with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$, and a polynomial bias field model were added to obtain the MR image $\boldsymbol{Y}$.

Now an extra step is added, where $\boldsymbol{Y}$ is not directly observed, but downsampled by a factor M resulting in a partial volumed MR image $\tilde{\boldsymbol{Y}} = \{\tilde{\boldsymbol{y}}_i,\ i{=}1, 2, \ldots I\}$ with only $I = J/M$ voxels (we assume here that no MR bias field is present in the images for the sake of simplicity). The downsampling process is illustrated in figure 23. The observed intensity $\tilde{\boldsymbol{y}}_i$ in voxel i of the downsampled image $\tilde{\boldsymbol{Y}}$ is modeled as the sum of the intensities $\boldsymbol{y}_j$ of all the subvoxels in the original image $\boldsymbol{Y}$ that underlie i. In voxels where not all subvoxels belong to the same tissue type, this causes partial voluming. Let $\boldsymbol{t}_i$ be a vector that contains the relative amount of each class k, k=1, 2, ... K in voxel i. A value of $t_{ik} = 1$ for some class k means that all the subvoxels underlying voxel i belong to class k, whereas a value of $t_{ik} = 0$ indicates that i does not contain class k at all. It can be shown that the observed intensity $\tilde{\boldsymbol{y}}_i$ in a voxel is governed by a normal distribution that only depends on its mixing proportions $\boldsymbol{t}_i$:

$$f(\tilde{\boldsymbol{y}}_i \mid \boldsymbol{t}_i, \Phi_Y) = G_{\tilde{\boldsymbol{\Sigma}}(\boldsymbol{t}_i)}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{\mu}}(\boldsymbol{t}_i))$$

with $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{t}_i) = M \cdot \sum_k t_{ik}\boldsymbol{\Sigma}_k$ and $\tilde{\boldsymbol{\mu}}(\boldsymbol{t}_i) = M \cdot \sum_k t_{ik}\boldsymbol{\mu}_k$.

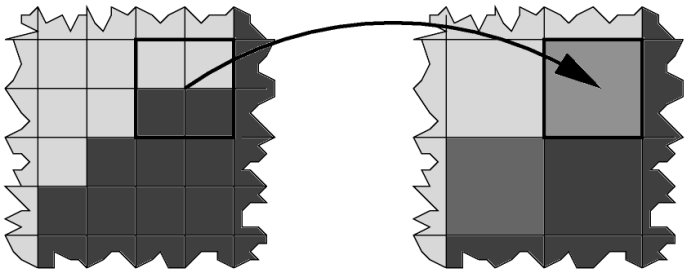


Figure 23: In the downsampling step, a number of voxels in the original image grid contribute to form the intensity of each voxel in the resulting image grid.

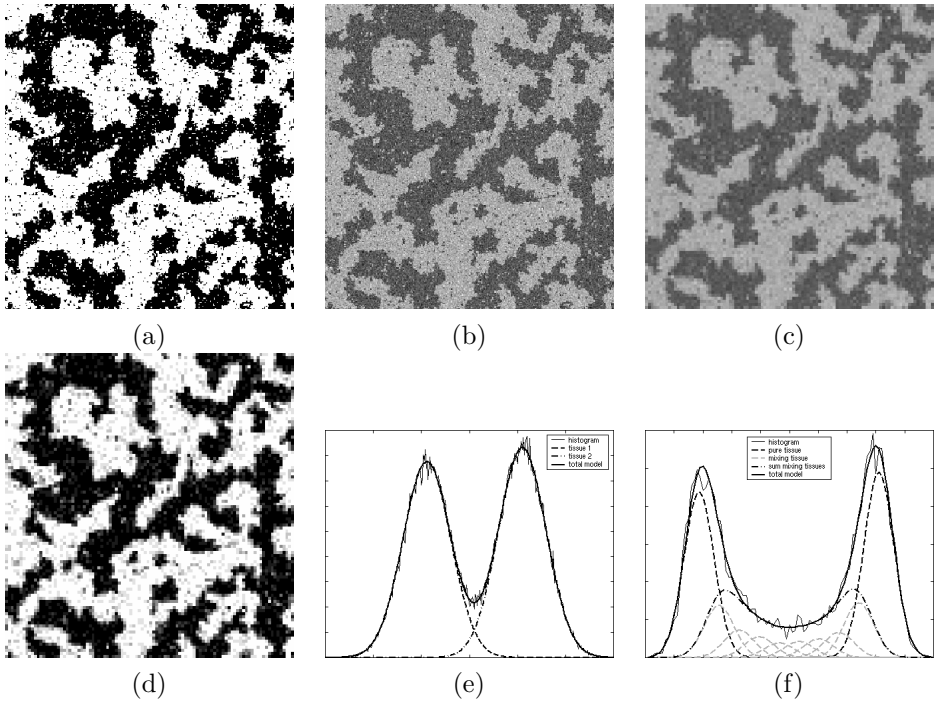


Figure 24: Illustration of the image model. First, a label image $\boldsymbol{L}$ is drawn according to some statistical model (a). An intensity image $\boldsymbol{Y}$ is obtained by adding tissue-specific normally distributed noise (b). Finally, Y is downsampled, resulting in an image $\tilde{\boldsymbol{Y}}$ that contains partial voluming (c). The underlying tissue fraction in $\tilde{\boldsymbol{Y}}$ is shown in (d), the histogram of $\boldsymbol{Y}$ along with its model in (e), and the histogram of $\tilde{\boldsymbol{Y}}$ with its model in (f). Note in subfigure (f) that the normal distributions associated with PV voxels do not all have the same weight. (Source: Publication IV)

Figure 24 illustrates this model on a 2-D example for two classes and $M = 3^2$ subvoxels per voxel. The effect of the downsampling is clearly visible in the histograms of $\boldsymbol{Y}$ and $\tilde{\boldsymbol{Y}}$ (figures 24 (e) and (f), respectively). Figure 24 (f) also shows the normal distributions $f(\tilde{\boldsymbol{y}}_i \mid \boldsymbol{t}_i, \Phi_Y)$ for each mixture $t = (\alpha, 1-\alpha)$ with $\alpha \in \{0, 1/M, 2/M, \ldots, (M-1)/M, 1\}$, weighted by the number of times each mixture occurs in the image. A model for the intensity distribution of PV voxels is obtained by summing the distributions of all non-pure mixtures.

7.2 Model parameter estimation

Because of the downsampling step, deriving an EM algorithm for estimating the model parameters from the data is somewhat more complicated than in the previous chapters. Previously, the image $\boldsymbol{Y}$ was directly observed, and the underlying tissue

labels $\boldsymbol{L}$ were unknown. This led to the EM strategy described in section 1.4, where the missing data $\boldsymbol{L}$ was iteratively filled in based on the current parameter estimates, and the model parameters were subsequently updated accordingly.

In contrast, now only a downsampled version of $\boldsymbol{Y}$ is available, and the original $\boldsymbol{Y}$ itself needs to be filled in by the EM algorithm as well. Therefore, the missing data that needs to be estimated in the expectation step is $\{\boldsymbol{L}, \boldsymbol{Y}\}$, which yields

Expectation step : find the function

$$Q(\Phi \mid \Phi^{(m-1)}) = E_{\boldsymbol{L},\boldsymbol{Y}}[\log f(\boldsymbol{Y}, \boldsymbol{L} \mid \Phi) \mid \tilde{\boldsymbol{Y}}, \Phi^{(m-1)}]$$

Maximization step: find

$$\Phi^{(m)} = \arg\max_{\Phi} Q(\Phi \mid \Phi^{(m-1)})$$

In Publication IV, we derived this algorithm for three different prior spatial models for the underlying label image $\boldsymbol{L}$:

Model A: no spatial correlation. Every mixing combination $\boldsymbol{t}$ has a spatially invariant prior probability $\pi_{\boldsymbol{t}}$.

Model B: no spatial correlation and uniform prior Every mixing combination $\boldsymbol{t}$ has a spatially invariant prior probability $\pi_{\boldsymbol{t}}$, that is the same for all non-pure $\boldsymbol{t}$.

Model C: Markov random field The spatial distribution of labels l_j in $\boldsymbol{L}$ is governed by an Ising/Potts MRF whose parameters regulate how much of each tissue is present and how voxels of a particular tissue type are clustered.

In each of these cases, the expectation step involves a statistical PV classification $f(\boldsymbol{t}_i \mid \tilde{\boldsymbol{Y}}, \Phi^{(m-1)})$ of the image voxels. For models A and B, this is given by

$$f(\boldsymbol{t}_i \mid \tilde{\boldsymbol{Y}}, \Phi^{(m-1)}) = \frac{f(\tilde{\boldsymbol{y}}_i \mid \boldsymbol{t}_i, \Phi_Y^{(m-1)}) \cdot \pi_{\boldsymbol{t}}^{(m-1)}}{\sum_{\boldsymbol{t}_i} f(\tilde{\boldsymbol{y}}_i \mid \boldsymbol{t}_i, \Phi_Y^{(m-1)}) \cdot \pi_{\boldsymbol{t}}^{(m-1)}}$$

which is clearly an extension of equation 4 used in section 1.4. Whereas previously, the voxels were classified into pure tissue types, voxels may now contain several tissues at the same time. Also the equations for $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are an extension of equations 5 and 6. In the special case where there is only $M = 1$ subvoxel per voxel, there is no downsampling, and the algorithm reduces to the original non-PV EM algorithm of section 1.4.

For model C, $f(\boldsymbol{t}_i \mid \tilde{\boldsymbol{Y}}, \Phi^{(m-1)})$ can not be calculated analytically. We therefore resort to the so-called MCEM algorithm [72] to approximate the expectation over the labels $\boldsymbol{L}$ by drawing a number of samples from the distribution $f(\boldsymbol{L} \mid \tilde{\boldsymbol{Y}}, \Phi^{(m-1)})$ by Monte Carlo simulation. This is a computationally very intensive and slow process, since it involves visiting each voxel hundreds of times per iteration of the EM algorithm.

7.3 Results

The method was implemented in 2-D, and the performance of each of the 3 different spatial models was validated on data that were simulated according to model C, i.e. where the underlying label image $\boldsymbol{L}$ is modeled as a MRF sample. Data were simulated according to different parameter sets Φ, and the three spatial models were fitted for each data set starting from the same randomized initial parameters. In all cases, the MRF of model C resulted in more accurate classifications than model A and B. Moreover, models A and B, which are entirely histogram-based, often failed to correctly estimate the underlying model parameters. In many cases the use of prior spatial knowledge as provided by model C showed indispensable for robust estimation of the model parameters.

Figure 25 shows an example with 2 classes and $M = 3^2$ subvoxels per voxel. Model A (figure 25 (d)) fits the histogram quite accurately, but the underlying model is not correctly estimated. Since there is no restriction on the weights $\pi_{\boldsymbol{t}}$ for the mixing proportions $\boldsymbol{t}$, the algorithm has simply adjusted these to get a good histogram fit, thereby setting the prior probability for pure tissue to zero. With model B (figure 25 (e)), all the mixing fractions corresponding to non-pure tissues are forced to have the same weight, resulting in the typical flat shape of the total intensity model for PV voxels that is commonly used in the literature [34, 73, 23, 74, 24]. However, the true mixing fractions in this example (figure 25 (b)) are not equal at all and therefore model B is condemned to fail. Only model C (figure 25 (f)) succeeds to retrieve the correct model parameters.

We also applied the algorithm to real MR brain images. Our current implementation of the MRF of model C is only 2-D, implying that tissue boundaries are assumed to be orthogonal to the image slice. Figure 26 (a) shows an axial slice of a high-resolution 1 mm isotropic T1-weighted image of the head through the central part of the brain where this assumption is more or less valid. Since different combinations of mixing tissue have overlapping intensities, model A is severely underconstrained in this case and was therefore not considered. Comparing the results obtained with model B and model C, it is clear that the MRF of model C reduces the noise in the segmentations considerably and forces partial volume voxels to lie on the border between the constituent tissues, in contrast to model B.

7.4 Discussion

The presented method provides a general framework for partial volume segmentation that encompasses and extends existing techniques. To our knowledge, no other technique has been described in the literature that enables estimation of tissue-specific means and covariance matrices guided by spatial information, while classifying the image voxels at the same time.

When model C is used, the method is an extension of the PV segmentation methods of Choi *et al.* [25], Pham and Prince [75] and Nocera and Gee [76]. Defining a MRF prior that imposes similar tissue combinations $\boldsymbol{t}_i$ over neighboring voxels, these methods iteratively assign a mixing fraction $\boldsymbol{t}_i$ to each voxel and update the

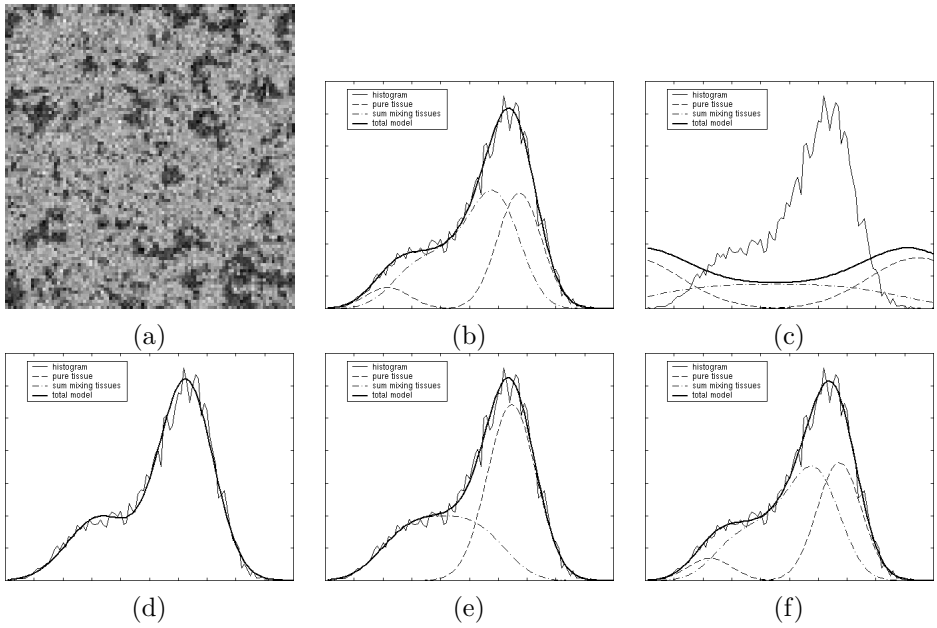


Figure 25: Performance of the 3 spatial models on simulated data with two tissues: simulated $\tilde{\boldsymbol{Y}}$ (a); histogram of $\tilde{\boldsymbol{Y}}$ with the underlying model overlayed (b); initialization for the model estimation (c); histogram fit with model A (d), with model B (e), and with model C (f). (Source: Publication IV)

mean intensity of every pure tissue type. However, it is assumed that all the tissues have the same diagonal covariance structure, which needs not be the case in real MR images as can be seen in for instance figure 22, and the problem of estimating this covariance is not addressed. Also, the voxel contents $\boldsymbol{t}_i$ are treated as model parameters that are estimated during each iteration using some greedy optimization algorithm. Referring to Titterington [77], this may introduce severe biases in the estimation of the mean intensities. In contrast, our algorithm also estimates tissue-specific covariance matrices, and all possible $\boldsymbol{t}_i$ are considered for the parameter estimation, contributing with a fraction $f(\boldsymbol{t}_i \mid \tilde{\boldsymbol{Y}}, \Phi)$. This ensures that no bias on the model parameters is incurred [77].

Choi *et al.* [25] and Nocera and Gee [76] used a MRF that simply imposes similar voxel contents $\boldsymbol{t}_i$ over neighboring voxels. While this model could easily be used in the framework presented here, we show in Publication IV that it is totally inappropriate as it encourages voxels to contain equal amounts of every tissue type everywhere in the image, in combination with extreme values for the mean intensities of pure tissue. Pham and Prince [75] used a MRF model that besides imposing similar voxel contents over neighboring voxels, explicitly favors pure tissue types in the voxels. However, they had to introduce a heuristic prior on the means to prevent them from taking extreme values, indicating that the MRF caveat might not have been totally solved. In contrast, our method defines an

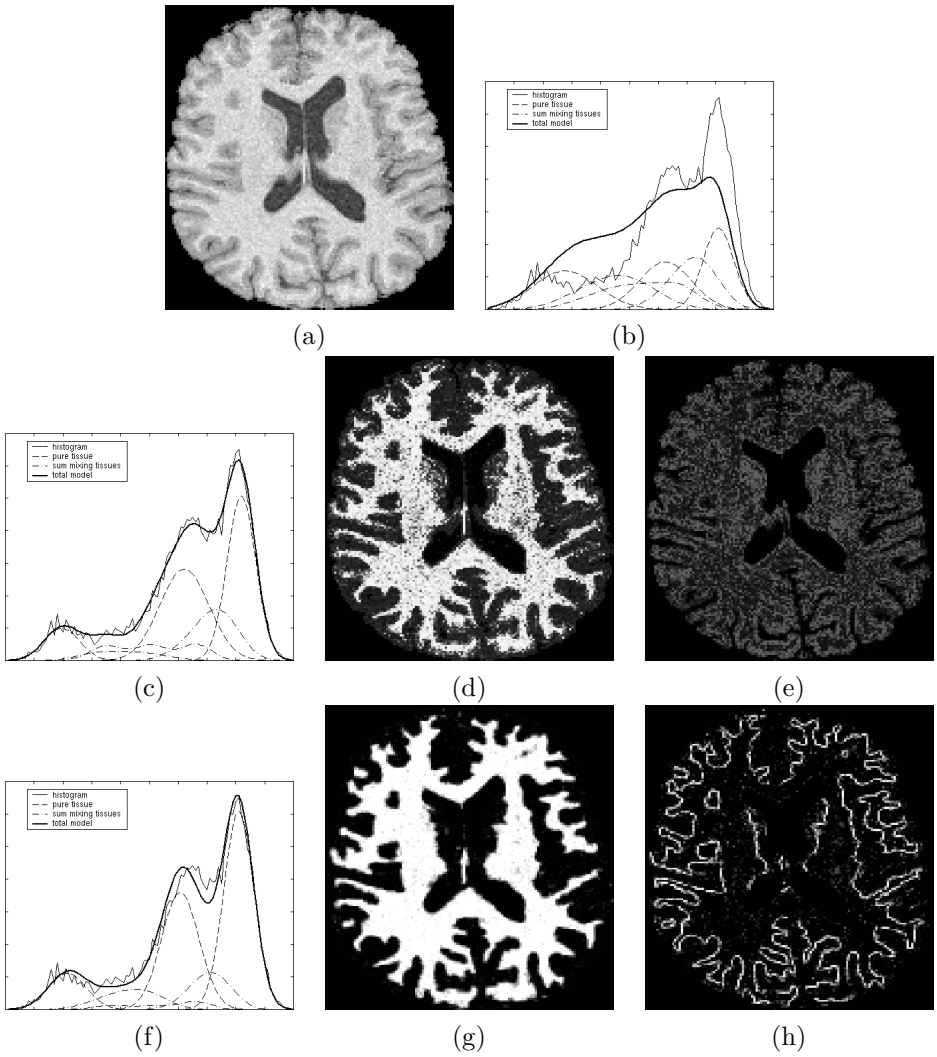


Figure 26: The advantage of a MRF prior illustrated on a sagittal slice of a high-resolution T1-weighted scan: image data $\tilde{\boldsymbol{Y}}$ (a); histogram with the initialization overlayed (b). Subfigures (c), (d) and (e) show respectively the histogram fit, the expected fraction of white matter and estimated probability for partial voluming between white matter-gray matter when model B was used. Subfigures (f), (g) and (h) show the same for model C. (Source: Publication IV)

Ising MRF model on subvoxels rather than directly on the voxels, thereby naturally imposing homogeneous regions of pure tissues bordered by PV voxels.

Our method is also an extension of the PV techniques of Santago and Gage [23, 74], Wu *et al.* [34], Laidlaw *et al.* [24] and Ruan *et al.* [73]. These methods estimate their parameters by minimizing the distance between the model and the histogram, thereby discarding all spatial information, and assume that when two tissues mix in a voxel, all mixing proportions are equally alike. When model B is used, our method forms an alternative fitting technique for these methods, optimizing the likelihood of the model instead of its distance to the histogram. However, the assumption of equally probable mixing fractions lacks any basis, as shown by Röll *et al.* [78] and as can also be seen from figures 24 (f) and 25 (b). Model A does not make any prior assumption about the mixing proportions at all, but this introduces so many degrees of freedom that the model fitting is severely underconstrained. Since our method optimizes the likelihood of the individual voxels rather than the distance to the histogram as in [34, 73, 23, 74, 24], spatial information derived from a prior spatial model can be used to guide the model fitting. As shown in section 7.3, this allows the mixing proportions to be non-uniform without making the estimation problem underconstrained.

With model C, our method iteratively interleaves a statistical classification of the image voxels using spatial information and an according update of the model parameters. In contrast, the methods in [34, 73, 24] first estimate the model parameters purely based on the histogram, and only subsequently introduce spatial information to classify the images. However, as shown in section 7.3, using the MRF *during* the parameter estimation itself is often indispensable to obtain reliable results. Because the model is so complex, there may be many parameter sets that all provide a close fit to the histogram. What discerns the true solution from the other ones is that it provides a meaningful classification in the images that corresponds to the spatial model. Therefore, spatial information is a prerequisite during the model fitting in such cases, and cannot be added afterwards to improve the segmentations.

We have come to similar conclusions when real lower-resolution MR images such as the ones shown in chapter 6 are processed. With model B, we always obtain close histogram fits, but we do not feel confident about the underlying parameters and classifications. We believe that inclusion of spatial information during the parameter estimation is necessary to robustly segment such images. However, we have not been able to apply model C on these data because our current 2-D implementation only models borders that are orthogonal to the image plane. We plan to implement model C in 3-D in the future so that partial voluming in the third dimension can also be taken into account. However, it is questionable if a stationary 3-D Ising/Potts MRF model is actually an appropriate description of the shape of the brain. In real MR data, large uniform regions belonging to one tissue type occur as shown in figure 26 (g), whereas such regions are seeded with many isolated voxels in samples of the MRF model as in figure 24 (a). The estimated MRF class transition costs are very high for real MR images, presumably because of the lack of such isolated voxels in the classifications. We expect that such high values for the MRF parameters will erase small details in the segmentations when lower-resolution images are processed,

because the histogram of such data sets contains more ambiguities. We believe that research to improved 3-D spatial models is the key to providing the indispensable prior information that is necessary for robust partial volume segmentation of lower-resolution MR images.

8 Conclusion

The aim of this thesis was to develop fully-automated methods for quantitative analysis of MR signal abnormalities in MS and CJD that can analyze large amounts of multi-spectral MR data in a reproducible way. The overall strategy adopted was to build statistical models for normal brain MR images, with stress on accurate intensity models. Signal abnormalities were detected as model outliers, i.e. voxels that could not be well explained by the model. Special attention was paid to automatically estimate all model parameters from the data itself, to eliminate subjective manual tuning and training.

8.1 Summary

In chapter 2, a method for fully automated bias field correction of MR images of the brain was presented. The tissue type of each voxel was assumed to be statistically independent from the tissue type of all other voxels. The observed MR images were modeled by adding tissue-specific normally distributed noise to each voxel, and additionally corrupting the images with a polynomial bias field model. The model parameters were estimated using an EM algorithm that iteratively improves the model parameters by interleaving three steps: classification of the voxels; estimation of the normal distributions; and estimation of the bias field. The bias field is estimated as a weighted least-squares fit to the difference between a predicted image without bias field and the actual image, with weights inversely proportional to the covariance of the tissue type to which each voxel is classified. It is therefore computed primarily from voxels that belong to classes with a narrow intensity distribution, such as white and gray matter, and extrapolated to regions where such estimate is ill-conditioned (CSF, non-brain tissues). In earlier approaches, tissue-specific intensity models were constructed by manually selecting representative points of each of the tissues considered, which needs to be repeated for each set of similar scans. In contrast, our method retrains such models on each individual scan being processed, starting from an initialization that is derived from a digital brain atlas that contains information about the a priori expected location of tissue classes. This allows full automation of the method without need for user interaction, yielding more objective and reproducible results.

Chapter 3 focused on the tissue segmentations generated by the algorithm. To discourage erroneous classifications of small regions surrounding the brain as gray matter or white matter, the model was extended by incorporating contextual information during classification using a MRF model. The MRF was designed to facilitate discrimination between brain and non-brain tissues while preserving the detailed interfaces between the various tissue classes within the brain. Since the exact calculation of the classification is not computationally feasible with the MRF model, a mean field approximation was adopted. The resulting EM algorithm interleaves 4 steps: classification of the voxels, estimation of the normal distribution parameters, estimation of the bias field, and estimation of the MRF parameters. The method was validated on simulated data, as well as on real MR data by com-

paring the automatic segmentations with manual delineations by a human expert. In both cases, the main source of errors was that partial voluming is not included in the model.

The method was further extended in chapter 4 in order to quantify MS lesions or CJD-related signal abnormalities in the images. Adding an explicit model for the pathological tissues is difficult because of the wide variety of their appearance in MR images, and because not every individual scan contains sufficient pathology for estimating the model parameters. These problems were circumvented by detecting lesions as voxels that are not well explained by the statistical model for normal brain MR images. Based on principles borrowed from the robust statistics literature, tissue-specific voxel weights were introduced that reflect the typicality of the voxels in each tissue type. These weights depend on an a priori chosen statistical significance level that defines when the belief that a voxel is an outlier exceeds the belief that it is a regular sample. Inclusion of these weights results in a robustized EM algorithm that simultaneously detects lesions as model outliers, and excludes these outliers from the model parameter estimation.

In chapter 5, this outlier detection scheme was applied for fully automatic segmentation of MS lesions from brain MR scans. To prevent partial volume voxels from being detected as model outliers as well, additional constraints on intensity were introduced in combination with a MRF that discourages voxels from being classified as MS lesion in the absence of neighboring white matter. The method is fully automated, with only one single parameter that needs to be experimentally tuned: a statistical significance threshold. Linear regression analysis between the TLL's estimated by the automatic algorithm and the TLL's derived from human expert segmentations on 20 data sets showed high correlation coefficients between 0.96 and 0.98 for a wide range of thresholds. Therefore, the actual choice of the threshold is fairly unimportant in clinical trials that assess changes in the TLL. However, it is much more critical when the goal is to investigate the basic MS mechanisms or time correlations of lesion groups in MS time series, as these applications require that the lesions are also spatially correctly detected. In general, the higher the resolution and the better the contrast between lesions and unaffected tissue in the images, the easier MS lesions are detected by the automatic algorithm and the higher the threshold should be chosen.

In chapter 6, application of the outlier detection technique did not show successful for automated quantification of the subtle MR intensity deviations in CJD. While abnormal signals were detected in the deep gray matter structures in some cases, considerable amounts of partial volume voxels also had high outlier belief values. A MRF approach to suppress these false positives as in MS was not feasible because the intensity deviations are much more subtle in CJD. It was therefore concluded that the partial volume effect needs to be explicitly included in the model.

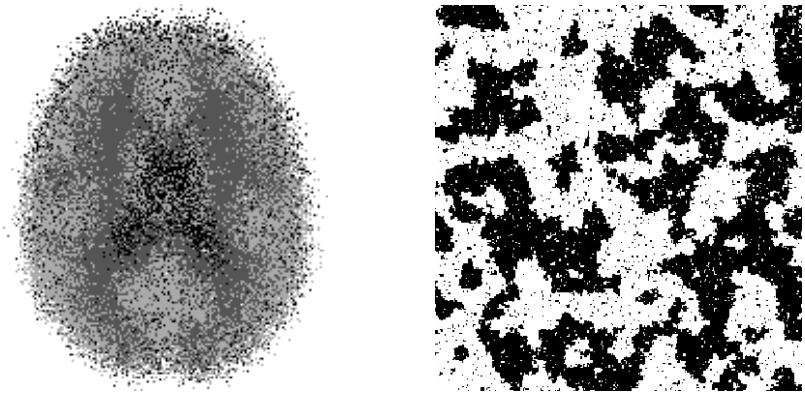
The image model was extended with a downsampling step in chapter 7, introducing partial voluming. An EM algorithm for estimating the model parameters was derived that includes the EM algorithm used in the previous chapters as a special case. It was demonstrated that the resulting method provides a general framework for partial volume segmentation that encompasses and extends existing techniques.

It was shown on simulated data that the histogram alone often does not provide enough information to uniquely define the underlying model parameters, making the use of appropriate prior spatial models during the parameter estimation indispensable. Experiments on 2-D slices of high-resolution MR scans demonstrated that a MRF prior reduces the noise in the segmentations considerably, and forces partial volume voxels to lie on the border between the constituent tissues. Without such a MRF, the parameter estimation was underconstrained for lower-resolution images, but the MRF approach could not be tested on such scans because our implementation did not take partial voluming out of the imaging plane into account. However, it was anticipated that a stationary MRF would erase small details in the segmentations when lower-resolution images are processed. It was therefore concluded that research to improved 3-D spatial models is necessary for general robust partial volume segmentation.

8.2 Segmentation methodology revisited

As discussed in section 1.3, geometry-driven and intensity-driven methods are the two main paradigms for model-based segmentation. In this thesis, the latter approach was adopted. Complex intensity models were developed that automatically fit to the data. As a result, multi-spectral MR data are segmented fully automatically without prior knowledge about the appearance of the different tissue types in the images. Bias fields are automatically corrected for, and the partial volume effect was explicitly taken into account. However, the efforts to describe the histogram of the images as accurately as possible contrast with the rather crude and primitive models for the spatial distribution of the tissue types in the images. It was simply assumed that the tissue type of each voxel is statistically independent from the tissue type of all other voxels, or only dependent on the nearest neighbors in the image grid. Figure 27 (a) shows a sample drawn from the digital brain atlas used throughout this thesis, assuming independence of the voxel labels. An example realization of the Ising/Potts MRF model is shown in figure 27 (b). While these models can be thought of as a representation of local random shape, they do not describe the global shape, symmetry properties and topology of the human brain.

In contrast, geometry-driven segmentation methods encode global shape and topology constraints, but heavily depend on prior assumptions about the intensities in the images. Image-based brain atlases have been used that deform to fit new images by minimizing voxel intensity differences, thereby directly transferring anatomical labels [79, 80]. While the deformations allowed by these methods are arbitrary, more powerful models have been constructed by statistical analysis of the contours/surfaces in a training set of individually segmented objects. In the Point Distribution Model (PDM) [81], a number of labeled key points are located on the object boundary and their position is subject to a principal component analysis (PCA), yielding a mean shape and a number of modes of variation around that mean shape. In a similar approach [82], the Fourier coefficients derived from the object contour were subject to the PCA instead of the key points. These statistical shape models restrict the possible shapes that can be retrieved from an image to a



(a) (b)

Figure 27: Samples from the spatial models used in this thesis: (a) a sample from the atlas shown in figure 9; (b) a sample from the Ising MRF model.

linear combination of modes of variation. While shape models have been used to segment the envelope of the brain and important sub-cortical structures, they typically have difficulty in segmenting such highly variable structures as the white-gray matter interface.

Intensity-driven and geometry-driven methods have traditionally been regarded as opposites. Attempts to combine the ability of intensity-driven methods to capture local shape variations with the general description of global anatomy provided by geometry-driven methods, have been limited to a subsequent use of both methods in separate processing steps [83, 22, 21, 84]. To our knowledge, no true statistical models for the shape of the human brain exist that both capture global shape and local variability. However, as intensity models become more complex with more parameters to be estimated from the data, there is a growing need for stronger shape models in intensity-driven methods to compensate for the extra degrees of freedom. In chapter 7, we argued that histogram fitting often becomes underconstrained for intensity models that take partial voluming into account. In order to robustly estimate the underlying model parameters, improved spatial models are needed that, in contrast to the models shown in figure 27, generate samples that effectively look like brains.

Recently, Neumann and Lorenz [85] proposed a powerful method for describing statistical shape that combines techniques from geometry-driven and intensity-driven methods. As in the PDM, a number of labeled key points are located on the object boundary, but their position is modeled by a MRF. Given the position of its so-called "neighbors", the position of a key point is conditionally independent from the position of all other key points. Since the "neighborhood" of a key point potentially contains both nearby and long-distant key points, this method bears the potential of combining the global shape description of PDMs with the necessary local flexibility.

The EM framework used throughout this thesis is general enough to allow inclusion of such improved spatial models in a natural way. This would result in a fusion of geometry-driven and intensity-driven methods, combining the advantages of both techniques. The possibilities of such a method are enormous, since it would allow fully automated partial volume segmentation and bias correction of multi-spectral MR images with unknown tissue contrast, while deforming a label atlas at the same time. The quantification of intensity abnormalities could be confined to anatomical regions of interest. The brain could be automatically segmented into relevant substructures, allowing the quantification of changes in shape and volume, over time in one individual patient, or between populations. Knowledge of the deformation of the label atlas would allow non-rigid multi-modal registration of images of different patients, and provide a common reference frame for population studies. Deriving realistic statistical models for the shape of the human brain is therefore a major challenge for further research.

References

[1] R.T. Johnson and C.J. Gibbs. Creutzfeldt-Jakob disease and related transmissible spongiform encephalopathies. *The New England Journal of Medicine*, 339(27):1994–2004, 1998.

[2] M. Zeidler, R.J. Sellar, D.A. Collie, R. Knight, G. Stewart, M.A. Macleod, J.W. Ironside, S. Cousens, A.F.C. Colchester, D.M. Hadley, and R.G. Will. The pulvinar sign on magnetic resonance imaging in variant Creutzfeldt-Jakob disease. *Lancet*, 355:1412–18, 2000.

[3] T.J. Crow. Functional specifications for morphological studies in schizophrenia. BIOMORPH deliverable fs1, 1996.

[4] A.C.F. Colchester. Functional specifications for multiple sclerosis, part 2. BIOMORPH deliverable fs2, November 1997.

[5] D.W. Paty, D.K.B. Li, the UBC MS/MRI Study Group, and the IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology*, 43:662–667, April 1993.

[6] R.I. Grossman and J.C. McGowan. Perspectives on multiple sclerosis. *AJNR*, 19:1251–1265, August 1998.

[7] J.H. Noseworthy, C. Lucchinetti, M. Rodriguez, and B.G. Weinshenker. Multiple sclerosis. *The New England Journal of Medicine*, 343(13):938–952, 2000.

[8] J.F. Kurtzke. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (edss). *Neurology*, 33:1444–1452, November 1983.

[9] A.C. Evans, J.A. Frank, J. Antel, and D.H. Miller. The role of MRI in clinical trials of multiple sclerosis: Comparison of image processing tech niques. *Annals of Neurology*, 41(1):125–132, january 1997.

[10] M. Filippi, M.A. Horsfield, P.S. Tofts, F. Barkhof, A.J. Thompson, and D.H. Miller. Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. *Brain*, 118:1601–1612, 1995.

[11] European project on brain morphometry (BIOMORPH, EU-BIOMED2 project nr. BMH4-CT96-0845, 1996–1998).

[12] J. Collinge. Variant Creutzfeldt-Jakob disease. *The Lancet*, 354:317–323, 1999.

[13] A.F. Hill, R.J. Butterworth, S. Joiner, G. Jackson, M.N. Rossor, D.J. Thomas, A. Frosh, N. Tolley, J.E. Bell, M. Spencer, A. King, S. Al-Sarraj, J.W. Ironside, P.L. Lantos, and J. Collinge. Investigation of variant Creutzfeldt-Jakob disease and other human prion diseases with tonsil biopsy samples. *The Lancet*, 353:183–189, 1999.

[14] A. Coulthard, K. Hall, P.T. English, P.G. Ince, D. J. Burn, and D. Bates. Quantitative analysis of MRI signal intensity in new variant Creutzfeldt-Jakob disease. *The British Journal of Radiology*, 72:742–748, August 1999.

[15] European project on quantitative analysis of MR scans in CJD (QAMRIC, EU-BIOMED2 project nr. BMH4-CT98-6048, 1998–2001).

[16] A. Zijdenbos, R. Forghani, and A. Evans. Automatic quantification of MS lesions in 3d MRI brain data sets: Validation of INSECT. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, volume 1496 of *Lecture Notes in Computer Science*, pages 439–448. Springer, 1998.

[17] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[18] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 2(1):1–36, 1996.

[19] J. Lötjönen, P.-J. Reissman, I.E. Mangin, and T. Katila. Model extraction from magnetic resonance volume data using the deformable pyramid. *Medical Image Analysis*, 3(4):387–406, 1999.

[20] X. Zeng, L.H. Staib, R.T. Schultz, and J.S. Duncan. Segmentation and measurement of the cortex from 3D MR images using coupled surfaces propagation. *IEEE Transactions on Medical Imaging*, 18(10):927–937, October 1999.

[21] M.A. González Ballester, A. Zisserman, and M. Brady. Segmentation and measurement of brain structures in MRI including confidence bounds. *Medical Image Analysis*, 4:189–200, 2000.

[22] C. Xu, D.L. Pham, M.E. Rettmann, D.N. Yu, and J.L. Prince. Reconstruction of the human cerebral cortex from magnetic resonance images. *IEEE Transactions on Medical Imaging*, 18(6):467–480, June 1999.

[23] P. Santago and H.D. Gage. Quantification of MR brain images by mixture density and partial volume modeling. *IEEE Transactions on Medical Imaging*, 12(3):566–574, September 1993.

[24] D. H. Laidlaw, K. W. Fleischer, and A. H. Barr. Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Transactions on Medical Imaging*, 17(1):74–86, february 1998.

[25] H. S. Choi, D. R. Haynor, and Y. Kim. Partial volume tissue classification of multichannel magnetic resonance images–a mixel model. *IEEE Transactions on Medical Imaging*, 10(3):395–407, september 1991.

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

[27] K. Held, E. R. Kops, B. J. Krause, W. M. Wells III, R. Kikinis, and H. W. Müller-Gärtner. Markov random field segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 16(6):878–886, december 1997.

[28] R. Guillemaud and M. Brady. Estimating the bias field of MR images. *IEEE Transactions on Medical Imaging*, 16(3):238–251, June 1997.

[29] S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Computer Science Workbench. Springer, 1995.

[30] Z. Liang, J. R. MacFall, and D. P. Harrington. Parameter estimation and tissue segmentation from multispectral MR images. *IEEE Transactions on Medical Imaging*, 13(3):441–449, September 1994.

[31] P. Schroeter, J.-M. Vesin, T. Langenberger, and R. Meuli. Robust parameter estimation of intensity distributions for brain magnetic resonance images. *IEEE Transactions on Medical Imaging*, 17(2):172–186, april 1998.

[32] D.L. Wilson and J.A. Noble. An adaptive segmentation algorithm for time-of-flight MRA data. *IEEE Transactions on Medical Imaging*, 18(10):938–945, october 1999.

[33] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.

[34] Z. Wu, H.-W. Chung, and F.W. Wehrli. A bayesian approach to subvoxel tissue classification in NMR microscopic images of trabecular bone. *MRM*, 31:302–308, 1994.

[35] A. Simmons, P. S. Tofts, G. J. Barker, and S. R. Arridge. Sources of intensity nonuniformity in spin echo images at 1.5 T. *Magnetic Resonance in Medicine*, 32:121–128, 1994.

[36] J. G. Sled and G. B. Pike. Understanding intensity non-uniformity in MRI. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'98*, volume 1496 of *Lecture Notes in Computer Science*, pages 614–622. Springer, 1998.

[37] M. Tincher, C.R. Meyer, R. Gupta, and D.M. Williams. Polynomial modeling and reduction of RF body coil spatial inhomogeneity in MRI. *IEEE Transactions on Medical Imaging*, 12(2):361–365, June 1993.

[38] S. E. Moyher, D. B. Vigneron, and S. J. Nelson. Surface coil MR imaging of the human brain with an analytic reception profile correction. *Journal of Magnetic Resonance Imaging*, 5(2):139–144, 1995.

[39] M. A. González Ballester. *Morphometric analysis of brain structures in MRI.* PhD thesis, Department of Engineering Science, University of Oxford, 1999.

[40] B. M. Dawant, A. P. Zijdenbos, and R. A. Margolin. Correction of intensity variations in MR images for computer-aided tissue classification. *IEEE Transactions on Medical Imaging*, 12(4):770–781, December 1993.

[41] C.R. Meyer, P.H. Bland, and J. Pipe. Retrospective correction of MRI amplitude inhomogeneities. In N. Ayache, editor, *Proc. First International Conference on Computer Vision, Virtual Reality, and Robotics in Medicine (CVRMED'95)*, volume 905 of *Lecture Notes in Computer Science*, pages 513–522. Springer, Nice, France, April 1995.

[42] A.C. Evans, D.L. Collins, S.R. Mills, E.D. Brown, R.L. Kelly, and T.M. Peters. 3d statistical neuroanatomical models from 305 MRI volumes. In *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, pages 1813–1817, 1993.

[43] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multi-modality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997.

[44] C. Brechbühler, G. Gerig, and G. Székely. Compensation of spatial inhomogeneity in MRI based on a parametric bias estimate. In *Proc. Visualization in Biomedical Computing (VBC '96)*, volume 1131 of *Lecture Notes in Computer Science*, pages 141–146. Springer, 1996.

[45] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A comparison of retrospective intensity non-uniformity correction methods for MRI. In *Proc. XVth International Conference on Information Processing in Medical Imaging (IPMI'97)*, volume 1230 of *Lecture Notes in Computer Science*, pages 459–464. Springer, 1997.

[46] M. Styner, C. Brechbühler, G. Székely, and G. Gerig. Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Transactions on Medical Imaging*, 19(3):153–165, March 2000.

[47] 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, February 1998.

[48] J.-F. Mangin. Entropy minimization for automatic correction of intensity nonuniformity. In *Proceedings of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis – MMBIA'00*, pages 162–169, 2000.

[49] B. Likar, M.A. Viergever, and F. Pernus. Retrospective correction of MR intensity inhomogeneity by information minimization. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*, volume 1935 of *Lecture Notes in Computer Science*, pages 375–384. Springer, 2000.

[50] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31:253–258, 1925.

[51] X. Descombes, J.-F. Mangin, E. Pechersky, and M. Sigelle. Fine structure preserving markov model for image processing. In *Proceedings of the 9th Scandinavian Conference on Image Analysis – SCIA 95*, pages 349–356, 1995.

[52] J. Zhang. The mean-field theory in EM procedures for markov random fields. *IEEE Transactions on Signal Processing*, 40(10):2570–2583, october 1992.

[53] D. A. Langan, K. J. Molnar, J. W. Modestino, and J. Zhang. Use of the mean-field approximation in an EM-based approach to unsupervised stochastic model-based image segmentation. In *Proceedings of ICASSP'92*, volume 3, pages 57–60, San Fransisco, CA, march 1992.

[54] R. K.-S. Kwan, A. C. Evans, and G. B. Pike. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging*, 18(11):1085–1097, November 1999. Available: http://www.bic.mni.mcgill.ca/brainweb/.

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

[56] A. P. Zijdenbos, B. M. Dawant, and R. A. Margolin. Intensity correction and its effect on measurement variability in the computer-aided analysis of MRI. In *Proc. 9th International Symposium and Exhibition on Computer Assisted Radiology (CAR'95)*, pages 216–221. Springer, Berlin, Germany, June 1995.

[57] W.J. Niessen, K.L. Vincken, J. Weickert, B.M. ter Haar Romeny, and M.A. Viergever. Multiscale segmentation of three-dimensional MR brain images. *International Journal of Computer Vision*, 31(2/3):185–202, 1999.

[58] P.J. Huber. *Robust Statistics*. Wiley series in probability and mathematical statistics. John Wiley and Sons, 1981.

[59] X. Zhuang, Y. Huang, K. Palaniappan, and Y. Zhao. Gaussian mixture density modeling, decomposition, and applications. *IEEE Transactions on Image Processing*, 5(9):1293–1302, september 1996.

[60] A.C.S. Chung and J.A. Noble. Statistical 3D vessel segmentation using a rician distribution. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'99*, volume 1679 of *Lecture Notes in Computer Science*, pages 82–89. Springer, 1999.

[61] D.C. Hoaglin, F. Mosteller, and J.W. Tukey, editors. *Understanding Robust and Explanatory Data Analysis*. Wiley series in probability and mathematical statistics. John Wiley and Sons, 1983.

[62] J.K. Udupa, L. Wei, S. Samarasekera, Y. Miki, M.A. van Buchem, and R.I. Grossman. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on Medical Imaging*, 16(5):598–609, october 1997.

[63] B. Johnston, M.S. Atkins, B. Mackiewich, and M. Anderson. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. *IEEE Transactions on Medical Imaging*, 15(2):154–169, april 1996.

[64] A. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer. Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Transactions on Medical Imaging*, 13(4):716–724, december 1994.

[65] M. Kamber, R. Shinghal, D.L. Collins, G.S. Francis, and A.C. Evans. Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images. *IEEE Transactions on Medical Imaging*, 14(3):442–453, september 1995.

[66] R. Kikinis, C.R.G. Guttmann, D. Metcalf, W.M. Wells III, G.J. Ettinger, H.L. Weiner, and F.A. Jolesz. Quantitative follow-up of patients with multiple sclerosis using MRI: Technical aspects. *Journal of Magnetic Resonance Imaging*, 9(4):519–530, April 1999.

[67] S. Warfield, J. Dengler, J. Zaers, C.R.G. Guttmann, W.M. Wells III, G.J. Ettinger, J. Hiller, and R. Kikinis. Automatic identification of grey matter structures from MRI to improve the segmentation of white matter lesions. *Journal Image Guided Surgery*, 1(6):326–338, 1995.

[68] A. Zijdenbos, A. Evans, F. Riahi, J. Sled, J. Chui, and V. Kollokian. Automatic quantification of multiple sclerosis lesion volume using stereotaxic space. In *Proceedings of Visualization in Biomedical Computing – VBC'96*, Lecture Notes in Computer Science, pages 439–448, 1996.

[69] C.R.G. Guttmann, R. Kikinis, M.C. Anderson, M. Jakab, S.K. Warfield, R.J. Killiany, H.L. Weiner, and F.A. Jolesz. Quantitative follow-up of patients with multiple sclerosis using MRI: Reproducibility. *Journal of Magnetic Resonance Imaging*, 9(4):509–518, April 1999.

[70] P. Demaerel, A.L. Baert, L. Vanopdenbosch, W. Robberecht, and R. Dom. Diffusion-weighted magnetic resonance imaging in Creutzfeldt-Jakob disease. *Lancet*, 349:847–848, March 1997.

[71] P. Demaerel, L. Heiner, W. Robberecht, R. Sciot, and G. Wilms. Diffusion-weighted MRI in sporadic Creutzfeldt-Jakob disease. *Neurology*, 52:205–8, January 1999.

[72] G. C. G. Wei and M. Tanner. A monte carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *J. Amer. Stat. Assoc.*, 85:699–704, 1990.

[73] S. Ruan, C. Jaggi, J. Xue, J. Fadili, and D. Bloyet. Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Transactions on Medical Imaging*, 19(12):1179–1187, December 2000.

[74] P. Santago and H. D. Gage. Statistical models of partial volume effect. *IEEE Transactions on Image Processing*, 4(11):1531–1540, november 1995.

[75] D. L. Pham and J. L. Prince. Unsupervised partial volume estimation in single-channel image data. In *Proceedings of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis – MMBIA'00*, pages 170–177, 2000.

[76] L. Nocera and J. C. Gee. Robust partial volume tissue classification of cerebral MRI scans. In K. M. Hanson, editor, *Proceedings of SPIE Medical Imaging 1997: Image Processing*, volume 3034 of *SPIE Proceedings*, pages 312–322. Bellingham, WA:SPIE, 1997.

[77] D.M. Titterington. Comments on "application of the conditional population-mixture model to image segmenation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(5):656–658, September 1984.

[78] S.A. Röll, A.C.F. Colchester, P.E. Summers, and L.D. Griffin. Intensity-based object extraction from 3D medical images including a correction for partial volume errors. In E. Hancock, editor, *Proc. BMVC '94*, pages 205–214, 1994.

[79] G.E. Christensen, S.C. Joshi, and M.I. Miller. Volumetric transformation of brain anatomy. *IEEE Transactions on Medical Imaging*, 16(6):864–877, December 1997.

[80] B.M. Dawant, S. L. Hartmann, J.-P. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel. Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part I, methodology and validation on normal subjects. *IEEE Transactions on Medical Imaging*, 18(10):909–916, October 1999.

[81] T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–366, July 1994.

[82] G. Székely, A. Kelemen, C. Brechbühler, and G. Gerig. Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible fourier contour and surface models. *Medical Image Analysis*, 1(1):19–34, 1996.

[83] F. Maes, K. Van Leemput, L. E. DeLisi, D. Vandermeulen, and P. Suetens. Quantification of cerebral grey and white matter asymmetry from MRI. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI'99*, volume 1679 of *Lecture Notes in Computer Science*, pages 348–357. Springer, 1999.

[84] T. Kapur, W.E.L. Grimson, W.M. Wells III, and R. Kikinis. Segmentation of brain tissue from magnetic resonance images. *Medical Image Analysis*, 1(2):109–127, 1996.

[85] A. Neumann and C. Lorenz. Statistical shape description using gaussian markov random fields and its application to medical image segmentation. In K. M. Hanson, editor, *Proceedings of SPIE Medical Imaging 1999: Image Processing*, volume 3661 of *SPIE Proceedings*, pages 192–203. Bellingham, WA:SPIE, 1999.

List of publications

International Journal

1. K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, P. Suetens: "Automated segmentation of multiple sclerosis lesions by model outlier detection", *IEEE transactions on medical imaging*, 2001 (accepted)

2. K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "Automated model-based tissue classification of MR images of the brain", *IEEE transactions on medical imaging*, vol. 18, no. 10, pp. 897-908, October 1999

3. K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "Automated model-based bias field correction of MR images of the brain", *IEEE transactions on medical imaging*, vol. 18, no. 10, pp. 885-896, October 1999

International Journal Abstract

1. K. Van Leemput, F. Maes, F. Bello, D. Vandermeulen, A. Colchester, P. Suetens: "Automated segmentation of MS lesions in MR", *NeuroImage*, vol. 11, part 2 of 2, no. 5, abstract no. S565, May 2000 (Proceedings 6th international conference on functional mapping of the human brain, June 12-16, 2000, San Antonio, Texas, USA)

2. L. Jyrkinen, K. Van Leemput, J. Oikarinen, O. Tervonen, O. Silven: "A method for patient movement registration during MRI intervention", *European Radiology*, vol. 7, pp. 1150-1151, 1997 (Proceedings 2nd interventional MRI symposium, October 1997)

International Conference: Published in Proceedings

1. S. Srivastava, K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "Validation of nonlinear spatial filtering to improve tissue segmentation of MR brain images", 4th international conference on medical image computing and computer-assisted intervention - MICCAI2001, October 14-17, 2001, Utrecht, The Netherlands (submitted)

2. K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "A statistical framework for partial volume segmentation", 4th international conference on medical image computing and computer-assisted intervention - MICCAI2001, October 14-17, 2001, Utrecht, The Netherlands (submitted)

3. F. Maes, K. Van Leemput, L.E. DeLisi, D. Vandermeulen, P. Suetens: "Quantification of cerebral grey and white matter asymmetry from MRI", Proceedings 2nd international conference on medical image computing and computer-assisted intervention - MICCAI'99, lecture notes in computer science, vol.

1679, pp. 348-357, September 19-22, 1999, Cambridge, UK, Taylor C. and Colchester A., eds., 1999, Springer

4. K. Van Leemput, F. Maes, F. Bello, D. Vandermeulen, A. Colchester, P. Suetens: "Automated segmentation of MS lesions from multi-channel MR images", Proceedings 2nd international conference on medical image computing and computer-assisted intervention - MICCAI'99, lecture notes in computer science, vol. 1679, pp. 11-21, September 19-22, 1999, Cambridge, UK, Taylor C. and Colchester A., eds., 1999, Springer

5. K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "Automatic segmentation of brain tissues and MR bias field correction using a digital brain atlas", Proceedings First international conference on medical image computing and computer-assisted intervention - MICCAI'98, lecture notes in computer science, vol. 1496, pp. 1222-1229, October 11-13, 1998, Cambridge, Massachusetts, USA

Internal Report

1. K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, P. Suetens: "Automated segmentation of multiple sclerosis lesions by model outlier detection", Internal report KUL/ESAT/PSI/0004, K.U.Leuven, ESAT, July 2000, Leuven, Belgium

2. K. Van Leemput, F. Maes, F. Bello, D. Vandermeulen, A. Colchester, P. Suetens: "Automated segmentation of MS lesions from multi-channel MR images", Internal report KUL/ESAT/PSI/9906, K.U.Leuven, ESAT, 1999, Leuven, Belgium

3. F. Maes, K. Van Leemput, L. DeLisi, D. Vandermeulen, P. Suetens: "Quantification of cerebral grey and white matter asymmetry from MRI" , Internal report KUL/ESAT/PSI/9905, K.U.Leuven, ESAT, April 1999, Leuven, Belgium

4. K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "Automated model-based tissue classification of MR images of the brain", Internal report KUL/ESAT/PSI/9903, K.U.Leuven, ESAT, March 1999, Leuven, Belgium

5. K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens: "Automated bias field correction and tissue classification of MR images of the brain using a digital atlas", Internal report KUL/ESAT/PSI/9901, K.U.Leuven, ESAT, February 1999, Leuven, Belgium

6. K. Van Leemput, D. Vandermeulen, P. Suetens: "Automatic segmentation of brain tissues and MR bias field correction using a digital brain atlas", Internal report KUL/ESAT/PSI/9806, K.U.Leuven, ESAT, June 1998, Leuven, Belgium

Book : Author of Part or Article

1. D. Vandermeulen, F. Maes, K. Van Leemput: "Quantitative MR imaging", Recent advances in diagnostic neuroradiology, pp. 47-64, Demaerel Ph., ed., 2000, Springer-Verlag Berlin Heidelberg