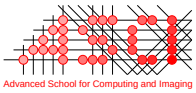
Automated Atlas-Based Segmentation of Brain Structures in MR Images

Application to a Population-Based Imaging Study

Fedde van der Lijn

Automated Atlas-Based Segmentation of Brain Structures in MR Images:
Application to a Population-Based Imaging Study
Fedde van der Lijn

The work described in this PhD thesis was conducted at the Departments of Radiology and Medical Informatics, Erasmus MC, the Netherlands, as a part of the ASCI graduate school.



This research was supported by the Netherlands Organisation for Scientific Research (NWO). Financial support for this thesis was kindly provided by the Department of Radiology, and the Erasmus University Rotterdam.

ISBN 978-90-8891-191-0
Printed by: Proefschriftmaken.nl | | Printyourthesis.com.

Automated Atlas-Based Segmentation of Brain Structures in MR Images

Application to a Population-Based Imaging Study

Automatische atlas-gebaseerde segmentatie van
breinstructuren in MR beelden
en het gebruik in een bevolkingsstudie

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit
Rotterdam op gezag van de rector magnificus, prof.dr. H.G. Schmidt,
en volgens het besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
Woensdag 8 september 2010 om 11:30 uur

door

Fedde van der Lijn

geboren op 5 februari 1977 te Eindhoven.

Promotor:	Prof.dr. W.J. Niessen
Overige leden:	Prof.dr. M.M.B. Breteler Prof.dr. P.J. Koudstaal Dr.ir. B.P.F. Lelieveldt
Co-promotor:	Dr.ir. H.A. Vrooman

Contents

Chapter 1

Introduction

The introduction of magnetic resonance imaging (MRI) has profoundly influenced neurology. In the 32 years since the first in-vivo brain scan was made[1], it has become an indispensable tool for studying disorders like schizophrenia [Wright et al., 2000], Alzheimer's disease [Jack et al., 1997] and Parkinson's disease [Laakso et al., 1996]. Structural MRI is used to visualize the brain changes underlying these diseases, functional MRI measures their effect on brain activity, and diffusion MRI can provide information on the brain's microstructural integrity. In short, MRI scanners have offered clinicians and researchers an unprecedented view inside the skull of their patients and subjects, and provided novel insights in determinants and clinical consequences of neurological diseases.

The adoption of MRI in the clinic has been relatively straightforward; the improved visualization of pathology was already a big leap forward compared to existing techniques like computed tomography (CT). As a result, the first case reports of cerebral MRI were already published in 1981 [Besson et al., 1981, Young et al., 1981]. The application of MRI in quantitative research has been more complicated, as it posed a new challenge: how to translate the information contained in an image into numbers? Researchers have dealt with this problem in a number of ways, notably by visual rating scales and manual segmentation protocols. In the first approach, an operator assigns a severity score to the pathology visible in an image. For example, white matter lesions can be rated with the Fazekas Scale [Fazekas et al., 1987]. Manual segmentation protocols are used as a guide to delineate specific brain structures in a consistent way. The resulting manual segmentations can then be used to extract quantitative biomarkers like the structure's volume. This approach was for example followed by Jack et al. [1997], who used the hippocampal segmentation protocol described in Jack [1994].

Automated image analysis methods provide a promising alternative for extracting quantitative information from neuroimaging data. They do not suffer from inter- and intrarater biases, and are able to process large amounts of data without human intervention. This last advantage is particularly important for large studies like the Rotterdam Study [Hofman et al., 2009] or the Alzheimer's Disease Neuroimaging Initiative (ADNI)[2] that collect MR images of several hundreds to thousands of subjects. Moreover, some types of analyses like the segmentation of the brain in gray and white matter, or cortical thickness measurement are so laborious that they are very rarely performed by operators.

In this thesis, methods for automated atlas-based brain structure segmentation are developed, evaluated, and applied to a large population neuroimaging study. This chapter puts these efforts in perspective by giving an overview of previous work in this field. Section 1.1 first describes the major automated brain structure segmentation paradigms, in particular atlas-based segmentation. Section 1.2 then presents some previous applications of automated segmentation to large neuroimaging studies. In section 1.3 the Rotterdam Scan Study is described, which is the source of all imaging data that was used in this work. Finally, the contributions

[1] nobelprize.org/nobel_prizes/medicine/laureates/1979/hounsfield-lecture.pdf
[2] www.adni-info.org

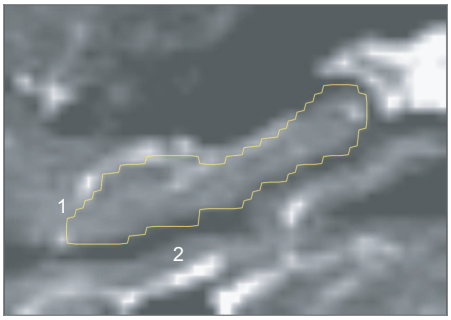


Figure 1.1: *A Half Fourier Acquisition Single Shot Turbo Spin Echo (HASTE) image of a human hippocampus (sagittal view) acquired with a 1.5 MRI unit. The yellow boundary marks the border of an expert segmentation. Note that the amygdala (1) and parahippocampal gyrus (2) have a similar intensity distribution.*

of this thesis are specified, and an overview of the following chapters is given.

1.1 Automated brain structure segmentation methods

Brain structure segmentation in MR images is a challenging task since the intensity distributions of different structures show considerable overlap. For example, the hippocampus has the same intensity values as the amygdala and the parahippocampal gyrus (Figure 1.1). Furthermore, not all boundaries are visible on MR images. This can also be appreciated from Figure 1.1, where a part of the border between the hippocampus and amygdala is invisible. Therefore, brain structure segmentation methods cannot rely on intensity information alone; they need additional knowledge to separate the structure of interest from the background.

The current brain structure segmentation literature can roughly be divided in three categories, based on the type of additional knowledge that they incorporate: deformable models, voxel classification, and atlas-based segmentation. In the following sections these paradigms will be briefly introduced. An overview of the work cited in this section can be found in Table 1.1. Of course, this categorization is not always clear-cut, as a combination of elements can boost a method's performance.

1.1.1 Deformable models

Methods based on the deformable model paradigm delineate brain structures by fitting a boundary model to the image, which incorporates some form of global shape knowledge. In snakes and level sets the shape information is relatively weak, as these techniques merely enforce smooth boundaries [Baillard et al., 2001, Ciofolo and Barillot, 2009, Shan et al., 2005].

Statistical shape models enforce stronger shape constraints and are therefore

Table 1.1: *A selection of published automated brain segmentation methods.*

	Method	Remarks
Deformable models	Kelemen et al. [1999]	Shape model; comb. w. atlas reg.
	Baillard et al. [2001]	Level set; comb. w. atlas reg.
	Joshi et al. [2002]	Shape model
	Shen et al. [2002]	Shape model
	Pizer et al. [2003]	Shape model
	Pitiot et al. [2004]	Shape model; comb w. atlas reg.
	Shan et al. [2005]	Level set; comb. w. atlas reg.
	Hu and Collins [2007]	Shape model; comb. w. atlas reg.
	Patenaude [2007]	Shape model; comb. w. atlas reg.
	Ciofolo and Barillot [2009]	Level set; comb. w. atlas reg.
Voxel classification	Arzhaeva et al. [2007]	
	Powell et al. [2008]	
	Morra et al. [2008]	
	Morra et al. [2010]	
Atlas-based segmentation	Collins et al. [1995]	
	Haller et al. [1997]	
	Iosifescu et al. [1997]	
	Dawant et al. [1999]	
	Fischl et al. [2002]	
	Hammers et al. [2003]	Average atlas
	Vemuri et al. [2003]	
	Carmichael et al. [2005]	
	Zhou and Rajapakse [2005]	Comb. with intensity model
	Barnes et al. [2007]	
	Heckemann et al. [2006]	Multi-atlas
	Pohl et al. [2006]	Comb. with intensity model
	Akselrod-Ballin et al. [2007]	Comb. with intensity model
	Gouttard et al. [2007]	Average atlas
	Hammers et al. [2007]	Comb. with intensity model
	Han and Fischl [2007]	Comb. with intensity model
	Pohl et al. [2007]	Comb. with intensity model
	Barnes et al. [2008]	Atlas selection
	Aljabar et al. [2009]	Atlas selection
	Chupin et al. [2009a]	Comb. with intensity model
	Wolz et al. [2010]	Atlas sel.; comb. with int. mod.

better equipped to deal with low-contrast boundaries. These models are constructed by parameterizing the shapes of manually labeled examples, and learning their mean shape and typical variations. Examples of shape models used for brain struc-

ture segmentation have been reported in Hu and Collins [2007], Joshi et al. [2002], Kelemen et al. [1999], Patenaude [2007], Pitiot et al. [2004], Pizer et al. [2003], Shen et al. [2002]. The most important difference between these methods is the flexibility of the parameterization. As a general rule, models with many degrees of freedom can describe a complex boundary, but also require more labeled examples to represent the potential shape variation.

The initialization of the boundary offers a way to incorporate spatial information in a deformable model. Deformable models are fit by minimizing a cost function that generally has several local minima. This usually ensures that the final result is close to the initialization and can therefore be used as a de-facto constraint on the spatial domain of the segmentation. In most of the methods cited above atlas-based segmentation is used to provide this initialization.

1.1.2 Voxel classification

Voxel classification methods segment brain structures voxel by voxel based on spatial and appearance information, which are represented by features. These features can include voxel coordinates describing the location, as well as intensity values of filtered versions of the MR image that describe the structure's appearance.

In classification-based segmentation image voxels are represented as points in a high-dimensional space, in which the coordinate axes are defined by the feature values. To segment an unlabeled target image, first manually labeled example images are used to train a voxel classifier. This is done by first sampling voxels from the training images and mapping them to the feature space. In the feature space a decision boundary is then found that best separates the voxels labeled as structure, from the voxels that were labeled as background in the examples. Different types of classifiers use different methods to derive this boundary from the training samples. After training, the classifier is applied to the unlabeled target voxels by mapping them in the feature space, and labeling them according to the decision boundary.

More features improve the classifier's ability to model the structure's appearance and location. However, it also tends to increase the complexity of the decision boundary which increases the risk of overtraining; the classifier is tuned too much on particularities of the training set, causing errors when classifying images that were not used for training. This risk can be decreased by increasing the number of examples, constraining the complexity of the decision boundary, or decreasing the number of features by removing those that are not very relevant for the classification accuracy.

The most important differences between brain structure segmentation methods based on voxel classification are the number and type of features used. In the work by Arzhaeva et al. [2007], Powell et al. [2008] up to 100 features were employed, including the intensities of voxels and their direct neighbours [Powell et al., 2008], or Gaussian scale space features [Arzhaeva et al., 2007]. Morra et al. [2008, 2010] segmented the hippocampus using thousands of features and the AdaBoost method [Freund and Schapire, 1997] that is robust against overtraining.

Throughout this thesis we use the word *atlas* for all manually labeled brain MR images that are used for segmentation based on registration. Traditionally, the term is reserved for a reference coordinate system that can be used to map information like the location of a brain structure or of brain activity in an fMRI experiment. This type of atlas can be based on a single brain (for example the well-known Talairach and Tournoux atlas [Talairach and Tournoux, 1988]), or it can be constructed from multiple brains (for example the ICBM atlases[3]). By using multiple brains it becomes possible to describe the anatomical variation of a population [Mazziotta et al., 1995].

1.1.3 Atlas-based segmentation

The final type of segmentation method is atlas-based segmentation (sometimes also called label propagation). In this approach, additional knowledge is introduced through an atlas image, in which an expert has labeled the brain structures of interest. The atlas is first registered to the target image, and the resulting transformation is then used to deform the atlas labels to the coordinate system of the target image[4]. During registration the similarity between the warped atlas image and the target image is maximized, while at the same time the deformation is constrained to ensure that the spatial information of the atlas is maintained.

Atlas-based segmentation approaches reported in the literature differ in the type of deformations they allow. Some authors have used affine [Carmichael et al., 2005] or piecewise affine [Barnes et al., 2007, Carmichael et al., 2005] registration algorithms. However, to account for the possible anatomical variation between the atlas and target image, typically a high-degree-of-freedom, non-rigid registration algorithm is employed [Collins et al., 1995, Dawant et al., 1999, Haller et al., 1997, Iosifescu et al., 1997, Vemuri et al., 2003] (for a comprehensive evaluation of the accuracy of atlas-based brain structure segmentation with several different publicly available registration methods see Klein et al. [2009]). More degrees of freedom means more flexibility to adapt the atlas to the target image, but also increases the risk that the optimization gets stuck in a local minimum that is far away from the optimal solution.

Since the similarity is computed over the entire image, atlas-based segmentations are prone to errors that do not have a large effect on this measure. Furthermore, the accuracy of the segmentation tends to decrease if the anatomy of the atlas is very different from the target. One possible solution to these problems is multi-atlas segmentation, in which several atlases are registered to the target image and the deformed labels are combined [Heckemann et al., 2006] (the idea was originally used to segment microscopy images of bee brains [Rohlfing et al., 2004]). The com-

[3] www.loni.ucla.edu/ICBM/Downloads/Downloads_Atlases.shtml

[4] Some of the papers cited in this section actually perform a segmentation by first registering the target image to the atlas image. The atlas labels can then be deformed to the target coordinate system by inverting the deformation. In the context of segmentation the result of both procedures are very similar, so we will not distinguish between them.

bination of multiple atlases greatly increases the accuracy and robustness of the segmentation. If a large number of atlases are available, results can even further be improved by selecting a subset of atlases that are very similar to the target [Aljabar et al., 2009, Barnes et al., 2008, Wolz et al., 2010].

Another way to increase the accuracy of atlas-based segmentation is by iteratively registering multiple manually labeled training images and averaging the deformations [Joshi et al., 2004]. The resulting deformations are then applied to both the training images and their manual labels. The result is a smoothed intensity image and a probability map which represents the average anatomical variation of the training population (these types of methods are also used to construct the reference atlases described in the boxed text). This average atlas can then be registered to an unlabeled target image [Gouttard et al., 2007, Hammers et al., 2003]. The approach is on average more accurate than registration with a single specific atlas, while requiring less computation than the multi-atlas approach (once the average atlas has been constructed). However, because it still relies on a single registration, average atlas segmentation is less accurate than registering multiple atlases Heckemann et al. [2006]. For this reason, information from a warped probability map is often combined with a classifier that describes the typical voxel intensities of the structure and its background [Akselrod-Ballin et al., 2007, Chupin et al., 2009a, Fischl et al., 2002, Hammers et al., 2007, Han and Fischl, 2007, Pohl et al., 2006, 2007, Wolz et al., 2010, Zhou and Rajapakse, 2005]. Throughout this thesis, examples of this method shall be referred to as *atlas&intensity-based*.

1.2 Automated segmentation in neuroimaging studies

Although the papers listed in Table 1.1 are just a selection of the work done on automated brain structure segmentation, most of these methods have not ventured outside the labs where they were developed. All of them have been validated with a limited number of images, but very few have actually been employed as an alternative to expert segmentations in neuroimaging studies. Some notable exceptions are listed in Table 1.2.

All three segmentation paradigms from Section 1.1 rely in some way on manually segmented training images, which must be similar to the unlabeled target image to produce accurate results. Therefore, if a researcher wishes to adopt an externally developed segmentation method, not only must he/she download or reimplement the software, but also create training data tailored to the intensity characteristics of his/her target images.

This has proved to be a major obstacle for the further dissemination of brain segmentation tools. The first five papers listed in Table 1.2 were all co-authored by people who were involved in the development of the segmentation methods that were used to perform the analyses. Only two methods are publicly available and have found use outside their "places of birth": FIRST and Freesurfer. Both techniques include an intensity normalization step to decrease the dependence on the specific characteristics of their training images.

Table 1.2: *Automated brain segmentation methods used in neuroimaging studies.*

	Method	Paradigm	Used in
Not publicly available	Haller et al. [1997]	atlas	Csernansky et al. [2005] and more
	Heckemann et al. [2006]	atlas	Heckemann et al. [2008] and more
	Powell et al. [2008]	mach. learn.	Ho and Magnotta [2010]
	Morra et al. [2008]	mach. learn.	Morra et al. [2009] and more
	Chupin et al. [2009a]	atlas	Chupin et al. [2009b]
Publicly available	Fischl et al. [2002]	atlas	Goldman et al. [2008] and more
	Patenaude [2007]	def. model	de Jong et al. [2008] and more
	various single atlas methods	atlas	Ikram et al. [2008] and more

FIRST is part of the FSL software library and is based on a statistical shape model [Patenaude, 2007]. It has so far been used in two small-sized studies of about 30 subjects [Péran et al., 2009, Seror et al., 2010], and three medium-sized studies of 70-140 persons [de Jong et al., 2008, Erickson et al., 2010, Janssen et al., 2009].

The brain structure segmentation method available in the Freesurfer package is part of a pipeline to measure and analyze cortical thickness. It is based on the method described in Fischl et al. [2002], in which the target image is affinely registered to an atlas that contains information on a structure's location, a location-specific intensity model, and the labels of its neighbors. This information is then combined in a Bayesian framework to find the maximum a posteriori solution. The version that is available for download contains some additional features compared to the original paper, including non-rigid registration of the atlas, and an intensity model that is derived from the MRI parameters [Fischl et al., 2004]. This last extension makes the method applicable to images that have different intensity characteristics than the training images.

An exhaustive overview of the use of Freesurfer is beyond of scope of this introduction, as it has been used extensively in neuroimaging studies. These include a 532-subject schizophrenia study [Goldman et al., 2008], and a large multi-center study of the influence of gender on the volume of brain structures that analyzed 1239 images [Fjell et al., 2009].

Apart from these two specific methods, several neuroimaging studies have used atlas-based segmentation, for example in DeCarli et al. [2005], Ikram et al. [2008], Wright et al. [2002]. The major advantage of atlas-based segmentation is that it is

easy to implement. Many registration techniques are publicly available and the method requires little training data; a single atlas can already produce relatively accurate segmentations. Furthermore, when a similarity measure like normalized mutual information is used, the method is relatively robust to intensity differences between the atlas and target images.

1.3 The Rotterdam Scan Study

The methods presented in this thesis are evaluated on and applied to a large number of MR images taken from the Rotterdam Scan Study. It is the neuroimaging component of the Rotterdam Study, a population-based cohort study on diseases among the elderly. The Rotterdam Study, which was started in 1990, follows more than 15,000 persons of 45 years and older who are living in the Ommoord neighborhood in Rotterdam [Hofman et al., 1991, 2007, 2009].

Every four years, participants are invited to the study center for a wide range of tests and examinations that focus on possible causes of cardiovascular, endocrine, hepatic, neurological, ophthalmic, psychiatric, and respiratory diseases. Furthermore, DNA is obtained from all participants and the researchers have access to health records of the participants' general practitioners. These efforts have led to a wealth of data that can be used to address a wide range of research questions. As of 2008, the findings of the Rotterdam Study have been presented in almost 1,000 research articles and reports [Hofman et al., 2009].

The Rotterdam Scan Study was set up within the Rotterdam Study to obtain more information on the causes and consequences of pre-symptomatic brain pathology. The first cohort of this study consisted of a random subset of Rotterdam Study participants between 60 and 90 years of age, who were invited to be scanned in 1995 and 1996. In the following years, these persons were invited for two rescans. Between 1999 and 2000, 244 subjects were scanned, and in 2006 images of 185 subjects were acquired using an upgraded scanner and imaging protocol. From 2005, the Rotterdam Scan Study was further expanded when an MRI scan with the recent protocol became part of the standard test routine for all Rotterdam Study participants.

The first Rotterdam Scan Study publications that involved brain structures were based on manual segmentations of the first cohort scanned in 1995 and 1996. Volumes measured in these segmentations were subsequently used in studies on the relation between hippocampal and amygdalar volume and APOE genotype, homocysteine, diabetes, blood pressure, and dementia [den Heijer et al., 2002, 2003a,b, 2005, 2006]. However, analysis of the follow-up images, the Scan Study expansion, or segmentation of any other structures than the hippocampus and the amygdala has not yet been conducted. Facilitating the segmentation of these images is an important application of the work presented in this thesis.

1.4 Thesis contributions and overview

The aim of this thesis is the development and application of accurate and robust brain structure segmentation methods for the automated analysis of large MR imaging studies. The main contributions of this work are the following:

1. The introduction of three methodological extensions to atlas-based segmentation (Section 1.4.1).
2. To assess whether automated segmentation can replace manual segmentation of the hippocampus in neuroimaging studies (Section 1.4.2).
3. An analysis of the risk factors of the rate of hippocampal atrophy, and its impact on cognition in a general elderly population (Section 1.4.3).
4. The introduction of an automated method to measure regional white matter lesion load to facilitate a more detailed study of its etiology and clinical consequences (Section 1.4.4).

The following sections will briefly discuss these contribution and provide an overview of the structure of this thesis.

1.4.1 Methodological extension of atlas-based segmentation

Brain structure segmentation methods that combine atlas registration with a statistical intensity model have been shown to produce very accurate results [Chupin et al., 2009a, Han and Fischl, 2007]. The work presented in this thesis builds on the success of this framework and extends it in three ways:

Global optimization of regularized atlas-based models

Many of the models used in atlas&intensity-based segmentation methods assume that voxels are independent. To reduce the effect of noise and produce smoother segmentations, some implementations include Markov Random Field (MRF) regularization [Chupin et al., 2009a, Fischl et al., 2002]. The optimizers used to solve these MRF models are not guaranteed to find a globally optimal solution. Graph cuts is a combinatorial optimization technique that can globally optimize MRF models and has therefore become very popular in computer vision [Boykov et al., 2001]. In Chapter 2, a regularized probabilistic atlas&intensity-based segmentation method is introduced that uses graph cuts. Also the other segmentation methods presented and used in this work are optimized with this technique.

Atlas&appearance-based segmentation

Several atlas&intensity-based methods rely on a global intensity model [Chupin et al., 2009a, Hammers et al., 2007, Pohl et al., 2006]. These models cannot adequately describe parts of the boundary where the background has a similar intensity as the structure of interest. As a result, atlas&intensity-based methods cannot

correct registration errors at these borders, nor can they segment structures with complex intensity patterns, like the cerebellum. In Chapter 3 multi-atlas-based segmentation is combined with an appearance model that is implemented with a classifier based on 65 Gaussian scale space features. This increases the applicability of the method, and also its robustness to registration errors.

Atlas&intensity-based measurement of the rate of hippocampal atrophy

The rate of hippocampal atrophy is an important indication of accumulating Alzheimer pathology. It can be measured from two sequential MR images by registering the baseline scan to the follow-up, and either quantifying changes of the resulting deformation field, or using it to deform a baseline segmentation [Barnes et al., 2008]. Just like atlas-based segmentation, this measurement is sensitive to errors that do not affect the global similarity measure driving the registration. In Chapters 4 and 5 the atlas&intensity-based method of Chapter 2 is adapted to measure the rate of hippocampal atrophy. The measurements are then used to study the relation between the atrophy rate, risk factors, and cognitive decline (see also Section 1.4.3).

1.4.2 Automated versus manual hippocampus segmentation in neuroimaging studies

As noted above, automated methods can potentially replace the laborious manual segmentation of brain structures, but are still little used in neuroimaging studies. All papers cited in Section 1.1 provide estimates of the segmentation accuracy, but very few analyze how this affects any associations computed with the automatically measured volumes. To assess this influence, the method presented in Chapter 2 is used to segment the hippocampus in the Rotterdam Scan Study cohort that was used in manual volumetry studies [den Heijer et al., 2002, 2003a, 2006]. Several of the previously reported associations are recomputed with the automatically measured volumes and compared.

1.4.3 Analysis of the hippocampal rate of atrophy

Previous work on the rate of hippocampal atrophy has shown its importance as a biomarker for development of dementia [Barnes et al., 2009]. But most of these studies were of limited size and follow-up time. In Chapter 4 an automated method is applied to measure the rate of hippocampal atrophy from the follow-up images of a Rotterdam Scan Study cohort. These data are then used to analyze the relation between hippocampal rate of atrophy and cognitive decline and dementia as measured during a ten-year follow-up period. In Chapter 5 we use the same measurements to study the association between APOE genotype, several cardiovascular risk factors, and hippocampal decline.

1.4.4 Automated measurement of regional white matter lesion volume

Elderly person often exhibit white matter lesions (WMLs), hyperintense foci on T2-weighted MR images, which are thought to play a role in cognitive decline [Frisoni et al., 2007]. Traditionally, lesions are divided in two categories: periventricular and subcortical WMLs, which are found adjacent to the ventricles and in the subcortical white matter respectively. Whether WMLs at different locations in the brain have different etiology and clinical consequences is unclear, however. Measurement of WML volume in smaller regions could provide more insight than this dichotomy. In Chapter 6, atlas-based segmentation techniques are used to develop a method that measures WML volume in regions based on their distance and orientation to the ventricles. Its potential is demonstrated by analyzing the relation between blood pressure and regional WML load in a Rotterdam Scan Study cohort.

References

A. Akselrod-Ballin, M. Galun, J. M. Gomori, A. Brandt, and R. Basri. Prior knowledge driven multiscale segmentation of brain MRI. In *MICCAI 2007, Lecture Notes in Computer Science*, volume 4792, pages 118–126, 2007.

P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage*, 46(3):726–738, 2009.

Y. Arzhaeva, E. van Rikxoort, and B. van Ginneken. Automated segmentation of caudate nucleus in MR brain images with voxel classification. In *3D Segmentation In The Clinic: A Grand Challenge*, pages 65–72, 2007.

C. Baillard, P. Hellier, and C. Barillot. Segmentation of brain 3D MR images using level sets and dense registration. *Med. Image Anal.*, 5(3):185–194, 2001.

J. Barnes, R. G. Boyes, E. B. Lewis, J. M. Schott, C. Frost, R. I. Scahill, and N. C. Fox. Automatic calculation of hippocampal atrophy rates using a hippocampal template and the boundary shift integral. *Neurobiol. Aging*, 28(11):1657–1663, 2007.

J. Barnes, J. Foster, R. G. Boyes, T. Pepple, E. K. Moore, J. M. Schott, C. Frost, R. I. Scahill, and N. C. Fox. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage*, 40(4):1655–1671, 2008.

J. Barnes, J. W. Bartlett, L. A. van de Pol, C. T. Loy, R. I. Scahill, C. Frost, P. Thompson, and N. C. Fox. A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. *Neurobiol. Aging*, 30(11):1711–1723, 2009.

J. A. Besson, A. I. Glen, E. I. Foreman, A. MacDonald, F. W. Smith, J. M. Hutchison, J. R. Mallard, and G. W. Ashcroft. Nuclear magnetic resonance observations in alcoholic cerebral disorder and the role of vasopressin. *Lancet*, 2(8252):923–924, 1981.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001.

O. T. Carmichael, H. A. Aizenstein, S. W. Davis, J. T. Becker, P. M. Thompson, C. C. Meltzer, and Y. Liu. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 27(4):979–990, 2005.

M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, O. Colliot, and A. D. N. Initiative. Fully automatic hippocampus segmentation and classification in alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus*, 19(6):579–587, 2009a.

M. Chupin, A. Hammers, R. S. N. Liu, O. Colliot, J. Burdett, E. Bardinet, J. S. Duncan, L. Garnero, and L. Lemieux. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *Neuroimage*, 46(3):749–761, 2009b.

C. Ciofolo and C. Barillot. Atlas-based segmentation of 3D cerebral structures with competitive level sets and fuzzy control. *Med. Image Anal.*, 13(3):456–470, 2009.

D. Collins, C. Holmes, T. Peters, and A. Evans. Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*, 3(3):190–208, 1995.

J. G. Csernansky, L. Wang, J. Swank, J. P. Miller, M. Gado, D. McKeel, M. I. Miller, and J. C. Morris. Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. *NeuroImage*, 25(3):783–792, 2005.

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 Trans. Med. Imaging*, 18 (10):909–916, 1999.

L. W. de Jong, K. van der Hiele, I. M. Veer, J. J. Houwing, R. G. J. Westendorp, E. L. E. M. Bollen, P. W. de Bruin, H. A. M. Middelkoop, M. A. van Buchem, and J. van der Grond. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain*, 131(Pt 12):3277–3285, 2008.

C. DeCarli, E. Fletcher, V. Ramey, D. Harvey, and W. J. Jagust. Anatomical mapping of white matter hyperintensities (WMH): exploring the relationships between periventricular WMH, deep WMH, and total WMH burden. *Stroke*, 36(1): 50–55, 2005.

T. den Heijer, M. Oudkerk, L. J. Launer, C. M. van Duijn, A. Hofman, and M. M. B. Breteler. Hippocampal, amygdalar, and global brain atrophy in different apolipoprotein E genotypes. *Neurology*, 59(5):746–748, 2002.

T. den Heijer, S. E. Vermeer, R. Clarke, M. Oudkerk, P. J. Koudstaal, A. Hofman, and M. M. B. Breteler. Homocysteine and brain atrophy on MRI of non-demented elderly. *Brain*, 126(Pt 1):170–175, 2003a.

T. den Heijer, S. E. Vermeer, E. J. van Dijk, N. D. Prins, P. J. Koudstaal, A. Hofman, and M. M. B. Breteler. Type 2 diabetes and atrophy of medial temporal lobe structures on brain MRI. *Diabetologia*, 46(12):1604–1610, 2003b.

T. den Heijer, L. J. Launer, N. D. Prins, E. J. van Dijk, S. E. Vermeer, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler. Association between blood pressure, white matter lesions, and atrophy of the medial temporal lobe. *Neurology*, 64(2):263–267, 2005.

T. den Heijer, M. I. Geerlings, F. E. Hoebeek, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler. Use of hippocampal and amygdalar volumes on magnetic resonance imaging to predict dementia in cognitively intact elderly people. *Arch. Gen. Psychiatry*, 63(1):57–62, 2006.

K. I. Erickson, R. S. Prakash, M. W. Voss, L. Chaddock, S. Heo, M. McLaren, B. D. Pence, S. A. Martin, V. J. Vieira, J. A. Woods, E. McAuley, and A. F. Kramer. Brain-derived neurotrophic factor is associated with age-related decline in hippocampal volume. *J. Neurosci.*, 30(15):5368–5375, 2010.

F. Fazekas, J. B. Chawluk, A. Alavi, H. I. Hurtig, and R. A. Zimmerman. MR signal abnormalities at 1.5 T in alzheimer's dementia and normal aging. *AJR Am. J. Roentgenol.*, 149(2):351–356, 1987.

B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

B. Fischl, D. H. Salat, A. J. W. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23 Suppl 1:S69–S84, 2004.

A. M. Fjell, L. T. Westlye, I. Amlien, T. Espeseth, I. Reinvang, N. Raz, I. Agartz, D. H. Salat, D. N. Greve, B. Fischl, A. M. Dale, and K. B. Walhovd. Minute effects of sex on the aging brain: a multisample magnetic resonance imaging study of healthy aging and Alzheimer's disease. *J. Neurosci.*, 29(27):8774–8783, 2009.

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997.

G. B. Frisoni, S. Galluzzi, L. Pantoni, and M. Filippi. The effect of white matter lesions on cognition in the elderly–small but detectable. *Nat. Clin. Pract. Neurol.*, 3(11):620–627, 2007.

A. L. Goldman, L. Pezawas, V. S. Mattay, B. Fischl, B. A. Verchinski, B. Zoltick, D. R. Weinberger, and A. Meyer-Lindenberg. Heritability of brain morphology related to schizophrenia: a large-scale automated magnetic resonance imaging segmentation study. *Biol. Psychiatry*, 63(5):475–483, 2008.

S. Gouttard, M. Styner, S. Joshi, R. G. Smith, H. Cody Hazlett, and G. Gerig. Subcortical structure segmentation using probabilistic atlas priors. In *SPIE Medical Imaging 2007: Image Processing*, volume 6512, 2007.

J. W. Haller, A. Banerjee, G. E. Christensen, M. Gado, S. Joshi, M. I. Miller, Y. Sheline, M. W. Vannier, and J. G. Csernansky. Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. *Radiology*, 202(2):504–510, 1997.

A. Hammers, R. Allom, M. J. Koepp, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, and J. S. Duncan. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*, 19(4):224–247, 2003.

A. Hammers, R. Heckemann, M. J. Koepp, J. S. Duncan, J. V. Hajnal, D. Rueckert, and P. Aljabar. Automatic detection and quantification of hippocampal atrophy on MRI in temporal lobe epilepsy: a proof-of-principle study. *NeuroImage*, 36(1): 38–47, 2007.

X. Han and B. Fischl. Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Trans. Med. Imaging*, 26(4):479–486, 2007.

R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.

R. A. Heckemann, A. Hammers, D. Rueckert, R. I. Aviv, C. J. Harvey, and J. V. Hajnal. Automatic volumetry on mr brain images can support diagnostic decision making. *BMC Med Imaging*, 8:9, 2008.

B.-C. Ho and V. Magnotta. Hippocampal volume deficits and shape deformities in young biological relatives of schizophrenia probands. *NeuroImage*, 49(4):3385–3393, 2010.

A. Hofman, D. E. Grobbee, P. T. de Jong, and F. A. van den Ouweland. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur. J. Epidemiol.*, 7(4):403–422, 1991.

A. Hofman, M. M. B. Breteler, C. M. van Duijn, G. P. Krestin, H. A. Pols, B. H. C. Stricker, H. Tiemeier, A. G. Uitterlinden, J. R. Vingerling, and J. C. M. Witteman. The Rotterdam Study: objectives and design update. *Eur. J. Epidemiol.*, 22(11): 819–829, 2007.

A. Hofman, M. M. B. Breteler, C. M. van Duijn, H. L. A. Janssen, G. P. Krestin, E. J. Kuipers, B. H. C. Stricker, H. Tiemeier, A. G. Uitterlinden, J. R. Vingerling, and J. C. M. Witteman. The Rotterdam Study: 2010 objectives and design update. *Eur. J. Epidemiol.*, 24(9):553–572, 2009.

S. Hu and D. L. Collins. Joint level-set shape modeling and appearance modeling for brain structure segmentation. *Neuroimage*, 36(3):672–683, 2007.

M. A. Ikram, H. A. Vrooman, M. W. Vernooij, F. van der Lijn, A. Hofman, A. van der Lugt, W. J. Niessen, and M. M. B. Breteler. Brain tissue volumes in the general elderly population. the Rotterdam Scan Study. *Neurobiol. Aging*, 29(6):882–890, 2008.

D. V. Iosifescu, M. E. Shenton, S. K. Warfield, R. Kikinis, J. Dengler, F. A. Jolesz, and R. W. McCarley. An automated registration algorithm for measuring MRI subcortical brain structures. *NeuroImage*, 6(1):13–25, 1997.

C. R. Jack. MRI-based hippocampal volume measurements in epilepsy. *Epilepsia*, 35 Suppl 6:S21–S29, 1994.

C. R. Jack, R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O'Brien, E. G. Tangalos, G. E. Smith, R. J. Ivnik, and E. Kokmen. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, 49(3):786–794, 1997.

J. Janssen, A. Diaz-Caneja, S. Reig, I. Bombín, M. Mayoral, M. Parellada, M. Graell, D. Moreno, A. Zabala, V. G. Vazquez, M. Desco, and C. Arango. Brain morphology and neurological soft signs in adolescents with first-episode psychosis. *Br. J. Psychiatry*, 195(3):227–233, 2009.

S. Joshi, S. Pizer, P. T. Fletcher, P. Yushkevich, A. Thall, and J. S. Marron. Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. *IEEE Trans. Med. Imaging*, 21(5):538–550, 2002.

S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23 Suppl 1:S151–S160, 2004.

A. Kelemen, G. Székely, and G. Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *IEEE Trans. Med. Imaging*, 18(10):828–839, 1999.

A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *NeuroImage*, 46(3):786–802, 2009.

M. P. Laakso, K. Partanen, P. Riekkinen, M. Lehtovirta, E. L. Helkala, M. Hallikainen, T. Hanninen, P. Vainio, and H. Soininen. Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia: An Mri study. *Neurology*, 46(3):678–681, 1996.

J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster. A probabilistic atlas of the human brain: theory and rationale for its development. the international consortium for brain mapping (icbm). *NeuroImage*, 2(2):89–101, 1995.

J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, X. Hua, A. W. Toga, C. R. Jack, M. W. Weiner, P. M. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Validation of a fully automated 3d hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *NeuroImage*, 43(1):59–68, 2008.

J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, A. W. Toga, C. R. Jack, N. Schuff, M. W. Weiner, P. M. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. *NeuroImage*, 45(1 Suppl):S3–15, 2009.

J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson. Comparison of adaboost and support vector machines for detecting alzheimer's disease through automated hippocampal segmentation. *IEEE Trans. Med. Imaging*, 29(1):30–43, 2010.

B. Patenaude. *Bayesian Statistical Models of Shape and Appearance for Subcortical Brain Segmentation*. PhD thesis, University of Oxford, 2007.

A. Pitiot, H. Delingette, P. M. Thompson, and N. Ayache. Expert knowledge-guided segmentation system for brain MRI. *NeuroImage*, 23 Suppl 1:S85–S96, 2004.

S. M. Pizer, P. T. Fletcher, S. C. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. A. Yushkevich, and E. L. Chaney. Deformable M-Reps for 3D medical image segmentation. *Int. J. Comput. Vis.*, 55(2-3):85–106, 2003.

K. M. Pohl, J. Fisher, W. E. L. Grimson, R. Kikinis, and W. M. Wells. A Bayesian model for joint segmentation and registration. *NeuroImage*, 31(1):228–239, 2006.

K. M. Pohl, S. Bouix, M. Nakamura, T. Rohlfing, R. W. McCarley, R. Kikinis, W. E. L. Grimson, M. E. Shenton, and W. M. Wells. A hierarchical algorithm for MR brain image parcellation. *IEEE Trans. Med. Imaging*, 26(9):1201–1212, 2007.

S. Powell, V. A. Magnotta, H. Johnson, V. K. Jammalamadaka, R. Pierson, and N. C. Andreasen. Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. *Neuroimage*, 39(1):238–247, 2008.

P. Péran, A. Cherubini, G. Luccichenti, G. Hagberg, J.-F. Démonet, O. Rascol, P. Celsis, C. Caltagirone, G. Spalletta, and U. Sabatini. Volume and iron content in basal ganglia and thalamus. *Hum Brain Mapp*, 30(8):2667–2675, 2009.

T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4):1428–1442, 2004.

I. Seror, H. Lee, O. S. Cohen, C. Hoffmann, and I. Prohovnik. Putaminal volume and diffusion in early familial Creutzfeldt-Jakob disease. *J. Neurol. Sci.*, 288(1-2): 129–134, 2010.

Z. Shan, Q. Ji, A. Gajjar, and W. Reddick. A knowledge-guided active contour method of segmentation of cerebella on MR images of pediatric patients with medulloblastoma. *J. Magn. Reson. Imaging*, 21(1):1–11, 2005.

D. Shen, S. Moffat, S. M. Resnick, and C. Davatzikos. Measuring size and shape of the hippocampus in MR images using a deformable shape model. *NeuroImage*, 15(2):422–434, 2002.

J. Talairach and P. Tournoux. *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging*. Thieme, 1988.

B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard. Image registration via level-set motion: applications to atlas-based segmentation. *Med. Image Anal.*, 7(1):1–20, 2003.

R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative. LEAP: learning embeddings for atlas propagation. *NeuroImage*, 49(2):1316–1325, 2010.

I. C. Wright, S. Rabe-Hesketh, P. W. Woodruff, A. S. David, R. M. Murray, and E. T. Bullmore. Meta-analysis of regional brain volumes in schizophrenia. *Am. J. Psychiatry*, 157(1):16–25, 2000.

I. C. Wright, P. Sham, R. M. Murray, D. R. Weinberger, and E. T. Bullmore. Genetic contributions to regional variability in human brain structure: methods and preliminary results. *NeuroImage*, 17(1):256–271, 2002.

I. R. Young, A. S. Hall, C. A. Pallis, N. J. Legg, G. M. Bydder, and R. E. Steiner. Nuclear magnetic resonance imaging of the brain in multiple sclerosis. *Lancet*, 2 (8255):1063–1066, 1981.

J. Zhou and J. C. Rajapakse. Segmentation of subcortical brain structures using fuzzy templates. *NeuroImage*, 28(4):915–924, 2005.

Chapter 2

Atlas- and intensity-based hippocampus segmentation

This chapter is based on:
Fedde van der Lijn, Tom den Heijer, Monique M.B. Breteler and Wiro J. Niessen. "Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts". *NeuroImage*, vol. 43, no. 4, pp. 708-720, 2008.

Abstract

In this chapter we present a segmentation method based on the minimization of an energy functional with intensity and prior terms, which are derived from manually labeled training images. The intensity energy is determined by a statistical intensity model that is learned from the training images. The prior energy consists of a spatial and regularity term. The spatial prior is obtained from a probability map created by registering the training images to the unlabeled target image, and deforming and averaging the training labels. The regularity prior energy encourages smooth segmentations. The resulting energy functional is globally minimized using graph cuts. The method was evaluated using image data from a population-based Rotterdam Scan Study. Two sets of images were used: a small set of 20 manually labeled magnetic resonance images and a larger set of 498 images, for which manual volume measurements were available, but no segmentations. This data was previously used in a volumetry study that found significant associations between hippocampal volume, cognitive decline, and the incidence of dementia. Cross-validation experiments with the labeled set showed similarity indices of 0.852 and 0.864, and mean surface distances of 0.40 and 0.36 mm for the left and right hippocampus. 83% of the automated segmentations of the large set were rated as 'good' by a trained observer. Also, the proposed method was used to repeat the manual hippocampal volumetry study. The automatically obtained hippocampal volumes showed significant associations with cognitive decline and dementia similar to the manually measured volumes. Finally, direct quantitative and qualitative comparisons showed that the proposed method outperforms a multi-atlas-based segmentation method.

2.1 Introduction

Many brain segmentation methods of magnetic resonance images (MRI) are based on the registration of one or more atlases [Collins et al., 1995, Dawant et al., 1999, Gouttard et al., 2007, Haller et al., 1997, Heckemann et al., 2006, Iosifescu et al., 1997, Vemuri et al., 2003]. The registration process is driven by a similarity measure that is computed over the entire image. As a result, atlas-based segmentations are prone to errors that do not have a large effect on this measure.

Adding a statistical intensity model can correct for these errors and increase the method's accuracy. Combined atlas and intensity models have been successfully applied to the segmentation of brain tissue [Ashburner and Friston, 2005, van Leemput et al., 1999] as well as brain structures [Akselrod-Ballin et al., 2007, Chupin et al., 2009, Fischl et al., 2002, Hammers et al., 2007, Han and Fischl, 2007, Pohl et al., 2006, 2007, Zhou and Rajapakse, 2005]. Most of these methods assume that voxels are independent of their neighbors, which makes them prone to noise.

Markov Random Field (MRF) regularization is a well-known technique to model the dependency between neighboring voxels, which can enforce more consistent labeling and reduces the effect of noise [Li, 1995]. Examples of brain segmentation methods that used MRF models can be found in Chupin et al. [2009], Fischl et al. [2002], Han and Fischl [2007], van Leemput et al. [1999]. However, these papers all solve these models with optimizers that are not guaranteed to find a globally optimal solution.

In this work we present a hippocampus segmentation method based on statistical models of intensity and spatial class label distributions, which are built from several labeled training images. The model is formulated as an energy functional, which is optimized using graph cuts [Boykov et al., 2001, Greig et al., 1989]. This combinatorial optimization technique is guaranteed to find a global minimum.

The energy functional contains two terms. The first is based on a voxel classifier that models the expected intensity distributions of the hippocampus and the surrounding tissue. These distributions are estimated from 20 training images of elderly subjects, in which the hippocampus was manually segmented by two trained observers. The second term models the expected spatial distribution of the hippocampus and background labels. It is derived from a regularizer and a probability map. The map is constructed from the labeled training brains after they have been registered to the target brain. This is comparable to an averaged, single-structure version of the atlas created with the multi-atlas segmentation method by Heckemann et al. [2006]. As shown in that work, multiple registrations are more accurate and robust than single atlas registration, used for example in Gouttard et al. [2007].

The method was evaluated using two sets of scans of elderly subjects from the Rotterdam Scan Study. The first set contained 20 manually labeled MR images. It was used to estimate the method's accuracy and to compute a quantitative comparison with a multi-atlas segmentation method. The second set consisted of 498 images for which manually measured volumes were available, but no segmentations. The automated segmentations of this set were visually rated and qualitatively

compared to the results of the multi-atlas segmentation.

Finally, to assess whether automated hippocampus segmentation can replace manual segmentation in neuroimaging studies, volumes were computed from set II. These were then used to compute several associations, which were compared to the associations found with the manual volumes. Some of the methods cited above have been applied in studies on the relation between automatically measured hippocampal volumes and neurological conditions [Csernansky et al., 1998, Goldman et al., 2008]. However, only in Pohl et al. [2007] the associations found with the automated method are directly compared with the associations found with manual volume measurements.

2.2 Method

Segmenting the hippocampus in an MR image can be compared to assigning a label $f_m \in \{0, 1\}$ to every voxel m in the image $\mathcal{M}$. Hippocampus voxels are designated by the label $f_m = 1$, background voxels by $f_m = 0$. As the segmentation is based on image intensities and prior knowledge, it can be naturally modeled as computing a maximum a posteriori (MAP) estimate using Bayes' theorem:

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{i}) = \frac{p(\mathbf{i}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{i})}, \tag{2.1}$$

in which $\mathbf{f}$ is the total label configuration and $\mathbf{i}$ contains all observed intensities of the image.

Finding the MAP estimate is equivalent to finding the minimum of an energy functional derived from equation 2.1. This functional is explained in more detail in the following section. The minimum energy is found using the graph cuts optimization method, which will be covered in section 2.2.2. Section 2.2.3 provides an overview of the method.

2.2.1 Energy model

By taking the negative logarithm of (2.1) and introducing a weighting parameter λ_1, segmentation becomes an energy minimization problem:

$$\begin{aligned} \hat{\mathbf{f}} &= \arg\min_{\mathbf{f}} \left(-\ln p(\mathbf{i}|\mathbf{f}) - \ln p(\mathbf{f}) \right) \\ &= \arg\min_{\mathbf{f}} \left(\lambda_1 E_{intensity}(\mathbf{f}) + E_{prior}(\mathbf{f}) \right). \end{aligned} \tag{2.2}$$

$p(\mathbf{i})$ does not depend on the labeling $\mathbf{f}$; it can therefore be ignored when optimizing the energy functional. The intensity energy $E_{intensity}$ is derived from the likelihood $p(\mathbf{i}|\mathbf{f})$ and models the observed intensities of the hippocampus and background classes. It measures how well a label configuration explains the image. The prior energy E_{prior} describes any prior knowledge on the class labels. In this work, it contains information on the spatial distribution of the labels and their neighbors, which is derived from registered training images.

Intensity energy

The intensity energy is based on a statistical voxel classifier and is a function of the total likelihood for all voxels. To estimate this function, it is assumed that the observed intensities are independent and identically distributed random variables. The image likelihood can then be written as a product of the likelihoods of the individual voxels m:

$$p(\mathbf{i}|\mathbf{f}) = \prod_{m \in \mathcal{M}} p(i_m|f_m). \tag{2.3}$$

The intensity energy is then given by:

$$E_{intensity}(\mathbf{f}) = -\sum_{m \in \mathcal{M}} \ln p(i_m|f_m). \tag{2.4}$$

The hippocampus consists predominantly of gray matter with relatively homogeneous intensities on MRI. The foreground likelihood $p(i_m|f_m = 1)$ can therefore be approximated by a Gaussian density function:

$$p(i_m|f_m = 1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{i_m - \mu}{\sigma}\right)^2\right]. \tag{2.5}$$

The mean μ and the variance σ are estimated with a training set of K_1 random samples drawn from J training images, in which the hippocampus was manually labeled. In all experiments, 250 foreground samples were taken per training image.

As the hippocampus borders on gray matter, white matter, and the temporal horn of the ventricles, the background distribution will be multimodal. The background likelihood $p(i_m|f_m = 0)$ is therefore approximated using a Parzen window estimator:

$$p(i_m|f_s = 0) = \frac{1}{\mathcal{K}_0} \sum_{k \in \mathcal{K}_0 1} \frac{1}{h} \phi\left(\frac{i_m - i_k}{h}\right), \tag{2.6}$$

where $\phi_i(i)$ is a Parzen kernel function with $\int \phi(i)di = 1$ and h is the kernel width. k indexes the $\mathcal{K}_0$ samples drawn from the voxels labeled as background in the training images. In this work, 250 background samples were taken from each training image. A Gaussian kernel function was used with a standard deviation of 30. Its width h was set equal to 121 (four times the standard deviation, plus one). These values were empirically chosen.

Prior energy

The prior energy is a function of the prior probability $p(\mathbf{f})$ of a particular label configuration $\mathbf{f}$. Modeling the joint probability for all voxels is undoable. Instead, we assume that the prior probability that voxel m has label f_m only depends on its position and its direct neighbors. $p(\mathbf{f})$ can therefore be approximated by a MRF with cliques consisting of one or two voxel sites [Li, 1995]:

$$p(\mathbf{f}) = \frac{1}{Z} \exp\left(-\sum_{m \in \mathcal{M}} \beta_m(f_m) - \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}_m} \beta_{m,n}(f_m, f_n)\right), \tag{2.7}$$

in which $\beta_m(f_m)$ and $\beta_{m,n}(f_m, f_n)$ are the one- and two-site clique potentials, and $\mathcal{N}_m \in \mathcal{M}$ is the set of neighbors of voxel m. Z is a normalization factor that does not depend on $\mathbf{f}$. It can therefore be ignored in the optimization.

Taking the negative logarithm gives the prior energy:

$$\begin{aligned} E_{prior}(\mathbf{f}) &= \sum_{m\in\mathcal{M}} \beta_m(f_m) + \sum_{m\in\mathcal{M}} \sum_{n\in\mathcal{N}_m} \beta_{m,n}(f_m, f_n) \\ &= E_{sprior}(\mathbf{f}) + E_{rprior}(\mathbf{f}). \end{aligned} \tag{2.8}$$

The prior energy consists of a spatial prior term E_{sprior}, determined by the one-site clique potentials, which models the expected spatial distribution of the class labels. The regularity prior energy E_{rprior} is proportional to the two-site clique potentials and is used to promote smooth segmentations.

The spatial prior energy is computed from a probability map built from the J manually labeled training images used to sample the intensity model. These training images are first non-rigidly registered to the target subject's coordinate system, and subsequently their labels are deformed and averaged. The map gives for every voxel location the probability of a foreground label $p(f_m = 1)$ and of a background label $p(f_m = 0) = 1 - p(f_m = 1)$. The spatial prior energy is then given by:

$$\beta_m(f_m) = -\lambda_2 \ln p(f_m), \tag{2.9}$$

with λ_2 a weighting parameter.

In this work, all registrations were computed in three steps: first the centers of mass were aligned, followed by an affine registration, and a non-rigid registration with a spline-based algorithm driven by mutual information [Rueckert et al., 1999]. The non-rigid registration was done in a multi-resolution fashion, using decreasing control point distances of 20, 10, 5, and 2.5mm.

As the hippocampus is a relatively smooth object, the following regularity prior energy is assumed:

$$\beta_{m,n}(f_m, f_n) = \begin{cases} 0 & \text{if } f_m = f_n \\ \frac{1}{2} B_{m,n} & \text{if } f_m \neq f_n \end{cases} \tag{2.10}$$

Following Boykov et al. [2001] we set the coefficient $B_{m,n}$ to

$$B_{m,n} = \frac{1}{d_{m,n}} \exp\left(-\frac{(i_m - i_n)^2}{2\zeta^2} \right), \tag{2.11}$$

where $d_{m,n}$ is the distance between voxel m and n. This gradient-modulated Ising model is commonly used in graph cut segmentation methods (for example in Boykov and Funka-Lea [2006], Freedman and Zhang [2005], Song et al. [2006]). $B_{m,n}$ penalizes different labels for neighboring voxels with an intensity difference of the order ζ. If the intensity jumps sharply between m and n, it is assumed that the voxels are of different classes, and only a small penalty is assigned. In this way, oversegmentation of sharp edges is prevented. ζ was chosen to be 200 in all experiments, which is approximately equal to the standard deviation of the foreground intensity distribution σ.

2.2.2 Optimization

The energy terms described in the previous two sections can be divided in two categories. The intensity energy and spatial prior energy functions both relate voxel m to a foreground or background model learned from training data. These terms are a function of a single label variable f_m. The regularity prior model relates two neighboring voxels, so it is a function of two label variables. The total energy functional is the sum of these terms for all voxels:

$$\begin{aligned} E(\mathbf{f}) &= \sum_{m \in \mathcal{M}} \lambda_1 E_{intensity}(f_m) + \lambda_2 E_{sprior}(f_m) + \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}_m} E_{rprior}(f_m, f_n) \\ &= \sum_{m \in \mathcal{M}} E^m(f_m) + \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}_m} E^{m,n}(f_m, f_n). \end{aligned} \tag{2.12}$$

Due to the between-voxel dependencies introduced by the two-variable terms, the energy model described in the previous sections is a complex functional of the segmentation $\mathbf{f}$. Local optimization methods like simulated annealing are not guaranteed to find a global minimum in polynomial time [Greig et al., 1989]. However, for Markov random field energy functionals of the form (2.12) it is possible to find the global optimum using graph cuts [Greig et al., 1989].

This optimization method works by converting the functional to a directional graph and computing the minimum s-t cut. The graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ has a set of nodes $\mathcal{V}$, one for every voxel in $\mathcal{M}$, plus two terminal nodes s and t (also called the source and sink). $\mathcal{E}$ is the set of edges connecting the nodes. Edges that connect a node to one of the terminals are called t-links and edges between two normal nodes, are known as n-links. Each of these edges has a non-negative weight assigned to it.

An s-t cut on $\mathcal{G}$ divides the nodes in two disjoint subsets S and T, such that $s \in S$ and $t \in T$. The cost of this s-t cut are the summed weights of all edges that go from S to T. If terminal t is regarded as the foreground class label and terminal s as the background class label, partitioning the nodes becomes equivalent to labeling the voxels. By choosing the edge weights based on energy functional (2.12), the cost of a cut is equal to the energy of the label configuration corresponding to that cut [Greig et al., 1989]. Finding the minimum energy then becomes equal to finding the minimum cut, which is a well-known problem in combinatorial optimization. Several methods exist that can find the globally minimum cut in polynomial time. We use the Maxflow algorithm introduced in Boykov and Kolmogorov [2004].

Figure 2.1 illustrates how the edge weights are determined in this work. All voxel nodes v_m are connected to both terminals. Consequently, for every node one of the t-links must be cut. As the t-link weights determine the cost of this cut, they represent the affinity between the terminals and the node. This is comparable to the one-variable energy terms in (2.12) that measure the "affinity" of the foreground and background model to voxel m. Detaching node v_m from the background terminal s is equivalent to giving voxel m a foreground label. The energy of this label configuration $f_m = 1$ is given by the intensity energy $E_{intensity}(f_m = 1)$ and the spatial prior energy $E_{sprior}(f_m = 1)$. The weight of the edge (s, v_m) that was cut is therefore set equal to $\lambda_1 E_{intensity}(f_m = 1) + \lambda_2 E_{sprior}(f_m = 1)$.

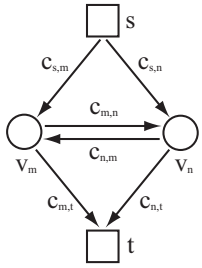


Figure 2.1: *The graph that represents the energy functional of a 1×2 voxel image. The edge weights have the following values:*
$c_{s,m} = \lambda_1 E_{intensity}(f_m = 1) + \lambda_2 E_{sprior}(f_m = 1)$
$c_{m,t} = \lambda_1 E_{intensity}(f_m = 0) + \lambda_2 E_{sprior}(f_m = 0)$
$c_{s,n} = \lambda_1 E_{intensity}(f_n = 1) + \lambda_2 E_{sprior}(f_n = 1)$
$c_{n,t} = \lambda_1 E_{intensity}(f_n = 0) + \lambda_2 E_{sprior}(f_n = 0)$
$c_{m,n} = E_{rprior}(f_m \neq f_n) + E_{rprior}(f_n \neq f_m)$
$c_{n,m} = E_{rprior}(f_n \neq f_m) + E_{rprior}(f_m \neq f_n)$

In a similar way the n-links represent the affinity between neighboring nodes and therefore correspond to the two-variable terms in (2.12). When both m and n have equal labels, the edges (v_m, v_n) and (v_n, v_m) do not form a connection between S and T and therefore do not have to be cut. But when $f_m \neq f_n$, v_m and v_n are connected to different terminals. One of the n-links forms a connection from source to sink that has to be cut at a cost equal to its weight. If both weights are set to $B_{m,n}/2$ the cut will correspond exactly to the regularizer of (2.10).

Not all energy functionals can be minimized with graph cuts. In Kolmogorov and Zabih [2004] the authors show that the minimum cut can only be computed in polynomial time if the energy functional is submodular, i.e.

$$\beta_{m,n}(0,0) + \beta_{m,n}(1,1) \leq \beta_{m,n}(1,0) + \beta_{m,n}(0,1). \tag{2.13}$$

For the model presented in this chapter, this clearly is the case.

2.2.3 Overview

In summary, the proposed method consists of the following steps. First, the three energy terms are constructed. The intensity energy model is trained by sampling from J labeled training images. The spatial prior energy model is derived from a probability map created in the unlabeled target image's coordinate frame. This is done by registering the training images to the target image. The probability map is then created by averaging the deformed training labels. The regularity prior energy is computed based on the intensities of neighboring voxel pairs in the target image. Next, a graph is constructed based on this energy functional and the minimum cut is computed. The optimal segmentation can then be created from the cut graph. Figure 2.2 shows an overview of these stages.

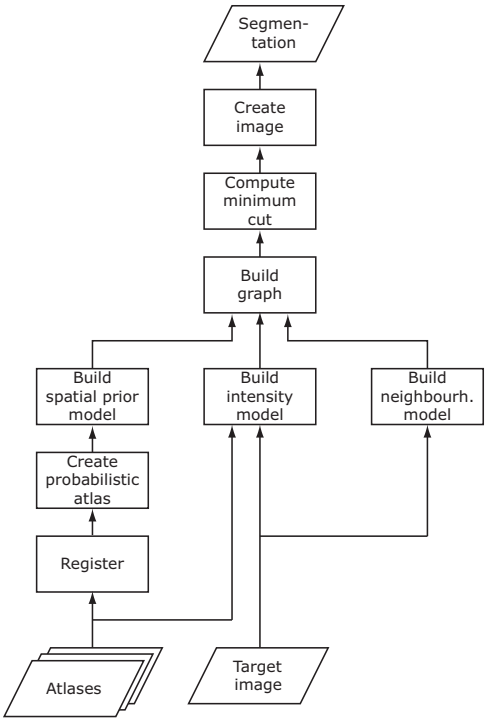


Figure 2.2: *Overview of the proposed graph-cut-based segmentation method.*

If the intensity and regularity terms are ignored, the graph-cut-based segmentation method will be equivalent to a multi-atlas segmentation method: a single-object version of the method presented in Heckemann et al. [2006]. In that case the optimal energy configuration would be achieved by labeling all voxels as foreground for which: $p(f_m = 1) > p(f_m = 0)$. The introduction of the intensity and regularity models influences the final segmentation at voxel locations around the boundary of the multi-atlas solution. In this region the intensity energy can compensate for possible registration errors. However, as the intensity energy functions are independent for all voxels, this can lead to segmentations with jagged edges. The regularizer will ensure this is less likely to happen.

2.3 Experiments and Results

The proposed graph-cut-based method was evaluated with two datasets: a small set of 20 images for which manual segmentations were available, and a large set of 498 unlabeled images. The image data and its preprocessing is described in more detail in the first section. Section 2.3.2 lists the quality measures used throughout this chapter. An intrarater study was performed, which is described in section 2.3.3. The experiments with image set I are covered in section 2.3.4. This set was used to estimate the graph-cut-based method's accuracy, and to compare it to the accuracy of a multi-atlas-based segmentation method. Furthermore, learning curves were computed for both methods by mapping their accuracy as a function of the number of training images used. In section 2.3.5 the proposed method was used to segment the images in set II. The results were evaluated by visual rating and qualitative comparison to the results of the multi-atlas method. Furthermore, volumes were derived from these segmentations and used to a perform a volume association study. The results were compared to the results of a manual volumetry study and the literature.

2.3.1 Image data

All experiments were performed using data from the Rotterdam Scan Study [den Heijer et al., 2003]. Five hundred eighteen elderly subjects of 55 years and older were scanned with a Siemens 1.5 T scanner, using an custom-designed, inversion recovery, 3D half-Fourier acquisition single-shot turbo spin echo sequence (HASTE). This sequence had the following characteristics: inversion time 4400 ms, repetition time 2800 ms, effective echo time 29 ms, matrix size 192×256, flip angle 180 degrees, slice thickness 1.25 mm, acquired in sagittal direction. The images were reconstructed to $128\times256\times256$ with a voxel dimension of $1.25\times1\times1$ mm. The images were corrected for inhomogeneities using N3 [Sled et al., 1998].

The image set was divided in two parts: set I contained images of 20 subjects selected to cover the population variation in age, sex, and hippocampus size. In these images the hippocampus was manually segmented by T.d.H. and another observer, who was trained by T.d.H. The segmentations of the second rater (25 % of the set) were revised and if necessary corrected by T.d.H.

The hippocampus formation included CA1 to CA4, the gyrus dentatus, and the subiculum. The images were segmented with the MNI Display visualisation and segmentation software [1], using a mouse-controlled cursor. The segmentations were performed on coronal slices, but with continuous reference to all orientations.

Set II contained all scans of the remaining 498 subjects. This data was used previously in several manual volumetry studies [den Heijer et al., 2002, 2003, 2006]. For this work, all images were manually segmented in 1999 and 2000 by T.d.H. and another observer, trained by T.d.H. Each rater processed approximately half of the images. In complicated cases the observers consulted each other. The seg-

[1] www.bic.mni.mcgill.ca/software/

Table 2.1: *Characteristics of the populations used for validation.*

	Set I (N = 20)	**Set II (N = 498)**
Mean age, years(SD)	74.6(8.2)	73.5(7.9)
Sex, % female	45	50
Manual segmentations available?	Yes	Volumes only
Incidence dementia, %*	0	9.6

* diagnosed during follow-up, see section 2.3.5

mentations were performed on a Siemens workstation using a mouse-controlled cursor. The rating was done on coronal slices perpendicular to the long axis of the hippocampus, without reference to other views. Subsequently, volumes were computed by counting the segmented voxels with an intensity in a predefined range that corresponded to gray matter.

The protocols of both segmentations were comparable, except for three differences. Firstly, for the segmentations of image set II, the posterior boundary of the hippocampus was defined as the coronal slice that contained the crux fornices in full profile. In image set I, the entire tail was included in the segmentation. Secondly, the segmentations of image set II were performed with reference only to coronal slices, instead of all three views. Thirdly, no intensity thresholds were used for the segmentations of set I. Therefore, any non-gray matter within the outlined region was included (for example small cysts).

At the time when the manual segmentation of image set II was performed, the authors were only interested in volumes and did not save the segmentations. Therefore, for this set only the measured volumes were available for our experiments, not the segmentations themselves. Some characteristics of both sets are listed in Table 2.1.

To speed up the registrations, a brain mask was created by non-rigidly registering a single manually segmented brain atlas to the target subject. The atlas did not include the cerebellum. This mask registration was performed with a B-spline-based method driven by mutual information [Klein et al., 2010]. The registered mask was also used to compute the intra-cranial volume (ICV), which was used to correct for head size differences in the volumetry study (see Ikram et al. [2008] for more details on the ICV measurement).

2.3.2 Quality measures

To compare a segmentation with a reference the following quality measures were used: Dice similarity index (SI), the relative volume difference (RV), the maximum and mean surface distance D_{max} and D_{mean}, and the volumetric intraclass correlation coefficient (ICC).

The SI is defined as:

$$SI = \frac{2V(\mathbf{f} \cap \mathbf{g})}{V(\mathbf{f}) + V(\mathbf{g})}, \tag{2.14}$$

where $V(\mathbf{f})$ and $V(\mathbf{g})$ are the volumes of the evaluated segmentation $\mathbf{f}$ and the reference $\mathbf{g}$.

The relative volume difference RV is given by:

$$RV = \frac{V(\mathbf{f}) - V(\mathbf{g})}{V(\mathbf{g})}, \tag{2.15}$$

The maximum surface distance is given by:

$$D_{max} = \max\left\{\delta(\mathbf{f}, \mathbf{g}), \delta(\mathbf{g}, \mathbf{f})\right\}, \tag{2.16}$$

with $\delta(\mathbf{f}, \mathbf{g})$ a set that contains the distances between every surface voxel in the evaluated segmentation $\mathbf{f}$, and the closest surface voxel in the reference $\mathbf{g}$.

The mean surface distance is defined by:

$$D_{mean} = \frac{\bar{\delta}(\mathbf{f}, \mathbf{g}) + \bar{\delta}(\mathbf{g}, \mathbf{f})}{2}, \tag{2.17}$$

with $\bar{\delta}(\mathbf{f}, \mathbf{g})$ the mean of set $\delta(\mathbf{f}, \mathbf{g})$ computed over all surface voxels of $\mathbf{f}$.

Finally, two-way random, absolute agreement, intraclass correlation coefficients (ICC) were computed between volumes derived from the segmentation and the reference [McGraw and Wong, 1996].

2.3.3 Intrarater reliability

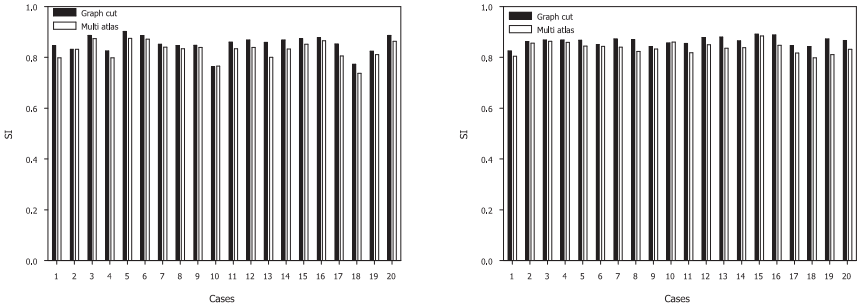
To assess the reproducibility of the manual segmentations of image set I, six images were segmented again by T.d.H. Time between the two segmentations was approximately two years. The results were evaluated by computing means, standard deviations and ranges for the Dice similarity index (SI), the relative volume difference (RV), and the maximum and mean surface distance D_{max} and D_{mean} with respect to the first segmentation.

The mean intrarater SI was 0.865±0.02 [0.835-0.898] for left and 0.865±0.02 [0.837-0.890] for right. D_{mean} was 0.33±0.08 [0.26-0.46] mm for the left and 0.33±0.08 [0.25-0.40] mm for the right side. The average D_{max} was 5.1 [3.4-7.5] and 4.8 [3.2-8.1] mm. The mean RV was -0.031±0.060 [-0.111-0.058] for left, and 0.00±0.034 [-0.049-0.045] for right. The ICCs were 0.77 and 0.89.

2.3.4 Experiments with ground truth

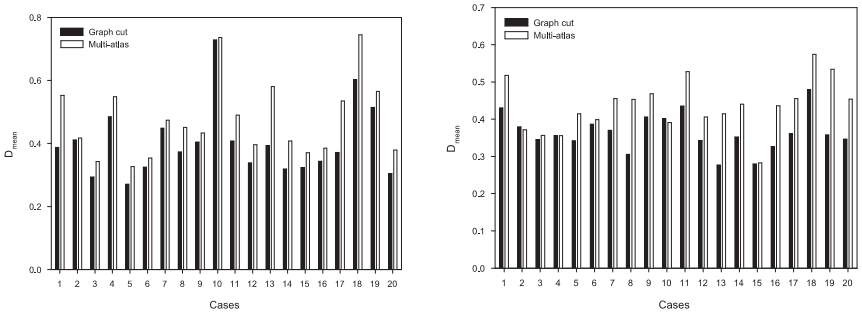
Parameter optimization and segmentation procedure

To segment the images in set I, each image was first registered to the others using the registration method and settings described in section 2.2.1. This resulted in 20 ×



(a) left hippocampus (b) right hippocampus

Figure 2.3: *Similarity indices of all images in set I for the graph cut and multi-atlas methods.*

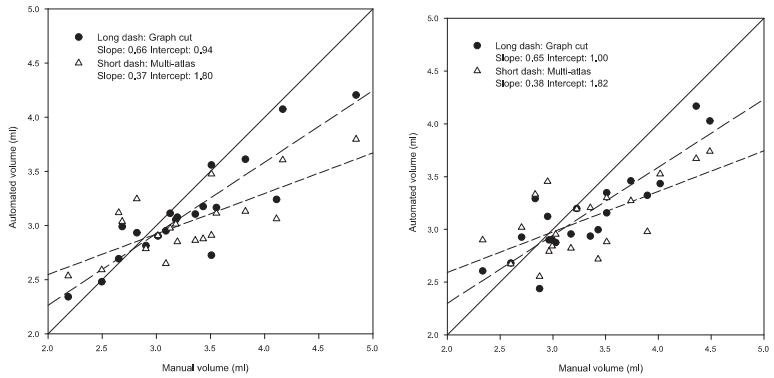


(a) left hippocampus (b) right hippocampus

Figure 2.4: *Mean surface voxel distances of all images in set I for the graph cut and multi-atlas methods.*

19 = 380 deformations. The segmentations were conducted in a leave-one-out fashion: one labeled image was used as target and the other 19 as training images. The manual segmentation of the target was used as the gold standard. This procedure was repeated until all 20 images in set I were segmented.

The values of λ_1 and λ_2 were chosen based on the labeled images that were used for training. This was done by performing leave-one-out parameter-tuning experiments on these 19 images. As an example, take the case that image 1 was used as a target, and images 2 to 20 were used for training. To determine the optimal values for λ_1 and λ_2, 19 parameter-tuning experiments were conducted. First, the manual labels of image 3-20 were warped to the coordinate framework of image 2, and a segmentation was computed using certain values of the weights. The manual labels of image 2 were then used to compute the SI for these weights. This was re-



(a) Manual vs automated volume for left hippocampus

(b) Manual vs automated volume for right hippocampus

Figure 2.5: *Manual and automated volumes for the graph cut and multi-atlas methods.*

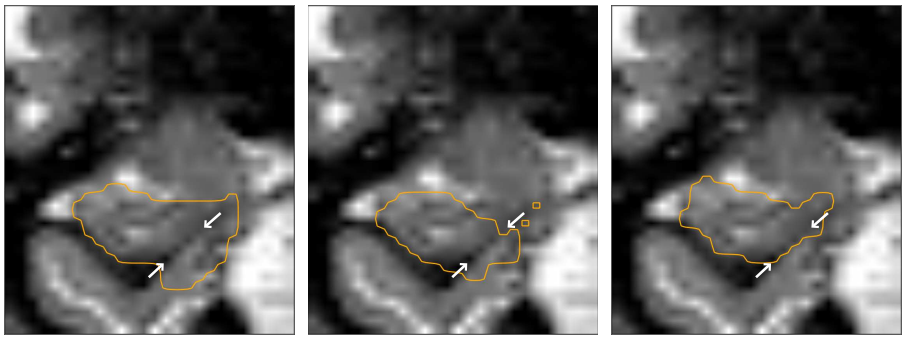
peated for several values of λ_1 and λ_2. Similar parameter-tuning experiments were then performed for images 3-20, each yielding a list of SI-weight combinations. The lambdas that gave the highest SI, averaged over all parameter-tuning experiments were chosen. Image 1 was then segmented by warping image 2-20 to its coordinate framework, creating a probability map, and cutting the graph, using these optimal weight values. In this way, the weights were chosen without using information of the target that would bias results.

To speed up the computation of the graph cut, subimages were taken around the left and right hippocampus for every subject. This was done by computing the union of the deformed training labels. A bounding box around this oversegmentation, expanded by two voxels in every direction, served as the region of interest. The typical size of these images was in the range of $30\times50\times30$ voxels.

Total computation time of the segmentation was dominated by the registration: each of the 19 registrations necessary to create the probability map took between 5 to 8 hours on a Linux server with a 2.2 GHz single core Opteron processor. Two machines with two processors were used to run these jobs. After a probability map was created, computing the graph cut took less than a second on a desktop computer with a 2.66 GHz Core processor.

Accuracy estimation

The proposed method's accuracy was evaluated by computing the Dice similarity index (SI), the relative volume difference (RV), and the maximum and mean surface distance D_{max} and D_{mean} with respect to the manual segmentation. Means, standard deviations and ranges of these scores were reported. ICCs were also computed to provide a measure of agreement between the manually and automatically



(a) Graph-cut-based hippocampus segmentation (b) Multi-atlas-based hippocampus segmentation (c) Manual hippocampus segmentation

Figure 2.6: *Example coronal slice showing the graph cut, multi-atlas, and manual segmentations of an oversegmented hippocampus. If the thin layer of white matter indicated by the arrows is ill-resolved, some of the training image registrations can cross over to neighboring gray matter structures (right and below of the white matter). The intensity model increases the error because it cannot distinguish between foreground and background gray matter.*

measured volumes.

The graph cut method's SI for every image in set I is shown in Figure 2.3 (black bars). The mean SI was 0.852±0.035 [0.763-0.902] for the left hippocampus and 0.864±0.016 [0.826-0.891] for the right. D_{mean} for all images is shown in Figure 2.4 (black bars). The average D_{mean} over all images was 0.40±0.11 mm [0.27-0.73 mm] and 0.36±0.05 mm [0.28-0.48 mm]. The average D_{max} was 4.8 mm [2.7-8.0 mm] on the left, and 4.6 mm [2.9-8.6 mm] on the right.

Figure 2.5 shows scatter plots of the graph-cut-estimated volumes versus the manually measured volumes for all target images (black circles). The RV was -0.042±0.083 [-0.223-0.114] and -0.040±0.090 [-0.151-0.162] for left and right. The volumetric ICCs were 0.82 for the left and 0.80 for the right side. The slope and intercept of the regression line fitted through the measurements in figure 2.5 were 0.66 (0.49-0.83 95%CI) and 0.94 ml (0.38-1.50 ml 95%CI) for left, and 0.65 (0.47-0.83 95%CI) and 1.00 ml (0.41-1.61 ml 95%CI) for the right hippocampus.

Visual inspection of the results showed that errors are mainly caused by undersegmentation occurring at the medial side of the hippocampal head and tail. These regions are not always completely covered by all deformed training labels. The intensity model of the proposed method can at least partially correct for these misregistrations. However, in some cases too many registration errors have accumulated, and the intensity model cannot completely compensate.

At other locations cases of oversegmentation were found. All results suffered from an oversegmentation at the anterior superior edge of the hippocampal head, where it borders the amygdala. However, this involved a very small number of voxels (20-50). In 5 hippocampi(12.5%) a part of neighboring gray matter areas like

the medial surface of the parahippocampal gyrus, or the collateral sulcus was labeled as foreground (Figure 2.6 shows an example). The hippocampus is separated from these areas by a layer of white matter that is not always well resolved. As a result, some of the training images are misregistered and cross over to the neighboring gray matter (Figure 2.6(b)). Since these regions have intensities similar to that of the hippocampus, the intensity model will then deteriorate the segmentation by pulling it further away from the actual boundary (Figure 2.6(a)).

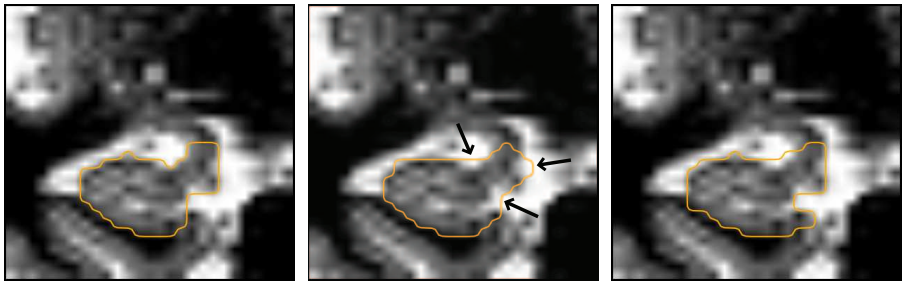
Quantitative comparison with multi-atlas segmentation

To compare these results with an established technique, image set I was also segmented with a multi-atlas-based method. This segmentation was obtained by assigning labels according to a 'winner takes all' principle. The resulting segmentation is equivalent to a single-object version of the method presented in Heckemann et al. [2006]. However, in that paper a probability map of multiple brain structures was created. Contrary to the version implemented here, this provides an explicit model of the background.

The SI results of the multi-atlas method are shown in Figure 2.3 and the mean distances in Figure 2.4(white bars). Compared to the graph cut method, its mean SIs were lower with 0.828$\pm$0.037 [0.737-0.875] for the left hippocampus, and 0.838$\pm$0.022 [0.798-0.884] for the right (versus 0.852 and 0.864). With scores of 0.47$\pm$0.12 mm [0.33-0.74mm] and 0.44$\pm$0.07 mm [0.28-0.57mm], D_{mean} was slightly higher for the multi-atlas method (0.40 and 0.39 for the graph cut method). All these differences were statistically significant with $p < 0.001$ in paired t-tests. The average D_{max} was comparable with 5.2 mm [2.6-9.0 mm] and 4.7 mm [2.5-8.1 mm] versus 4.8 mm and 4.7 mm.

The multi-atlas method's RV was -0.059$\pm$0.130 [-0.255-0.176] for the left hippocampus, and -0.047$\pm$0.133 [-0.235-0.241] for the right. This did not differ significantly from the RV scores of the graph-cut-based method (paired t-test: $p = 0.36$ and $p = 0.64$). However, the standard deviation and range of the multi-atlas were much larger. This is also reflected in the slope and intercept of the regression line fitted through the data in Figure 2.5: 0.37 (0.20-0.55 95%CI) and 1.80 ml (1.22-2.37 ml 95%CI) for left, and 0.38 (0.16-0.61 95%CI) and 1.82 ml (1.08-2.56 ml 95%CI) for the right hippocampus. The volumes measured by the proposed method are closer to the manual volumes than the volumes measured with the multi-atlas method. Volume differences in the population are reduced when measured with both automated techniques, but this effect is stronger for the multi-atlas-based method. The multi-atlas ICC was much lower than the ICC of the graph cut method: 0.51 for both sides, versus 0.82 and 0.80.

An example slice of the results of the graph cut and multi-atlas segmentation is shown together with the manual segmentation in Figure 2.7. Visual comparison of the results show that the graph-cut-based method corrects for registration errors at voxels where the foreground or background intensities clearly deviate from the intensity model.



(a) Graph-cut-based hippocampus segmentation (b) Multi-atlas-based hippocampus segmentation (c) Manual hippocampus segmentation

Figure 2.7: *Example coronal slice showing the graph cut, multi-atlas, and manual segmentations. The intensity model of the graph-cut-based method corrects some of the oversegmentations of cerebrospinal fluid that can be seen in the multi-atlas segmentation (marked by the arrows).*

Learning curve estimation

To investigate the relation between the accuracy and the number of training images used, a collection of probability maps was created. Each probability map $PA(J,t)$ is characterized by t, the coordinate framework of the target image in which the map was built, and J, the number of training images used to construct it. J was chosen to be an odd number from 3 to 19. For every value of J below 19, 100 unbiased probability maps were created: 5 per target t. This was done by randomly selecting a training images from image set I, excluding the target image t. For $J = 19$, only one map could be generated per target. Matching intensity training sets were constructed by drawing 1500 samples per training image, per side.

From these probability maps, segmentations were computed using the graph cut and multi-atlas methods. The parameters of the graph-cut-based segmentation were fixed for all experiments. The results were compared to the manual labels of target t by computing SIs, and averaging these over all 100 atlases and over the left and right side.

Figure 2.8 shows the mean SI as a function of J. The graph-cut-based method has a higher mean and smaller ranges than the multi-atlas-based method for every J. All SI differences between the two methods are statistically significant with a $p < 0.001$ in paired t-tests. This difference decreases with the number of training images: for $J = 3$, the mean SI was 0.825 for the graph-cut-based method and 0.793 for the multi-atlas method. For $J = 19$, the mean SIs were 0.853 and 0.835. The difference with the mean SIs estimated in the accuracy experiment and the quantitative comparison is due to different weight settings.

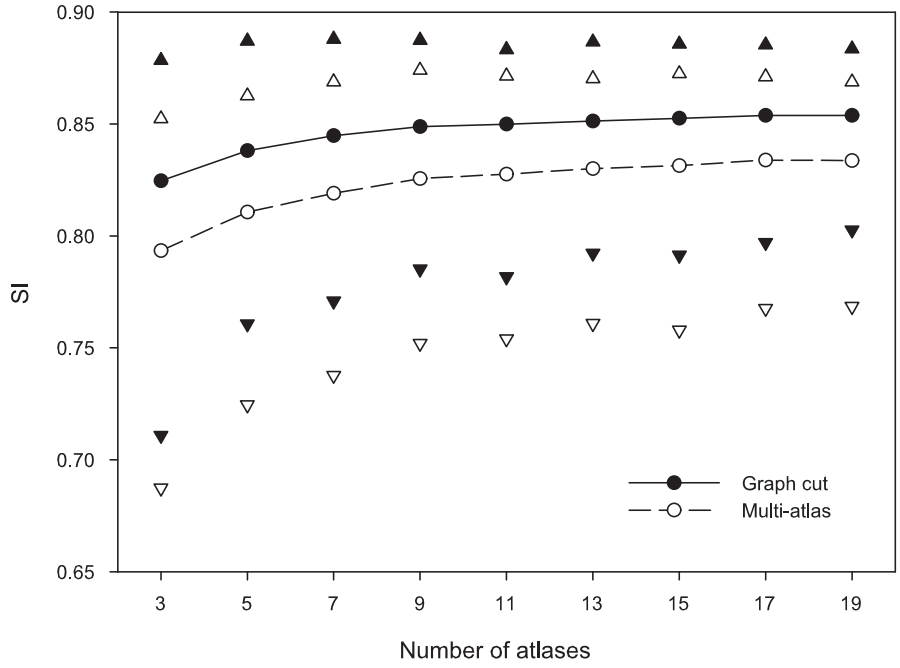


Figure 2.8: *The similarity index as a function of the number of training images used for the multi-atlas and graph-cut methods. The triangles indicate minimum and maximum values.*

2.3.5 Experiments without ground truth

Parameter optimization and segmentation procedure

The graph cut method's performance on large sets of images was assessed by applying it on image set II, using set I as training images. Every training image was registered to the target image using the method described in section 2.2.1. The 498 $\times$ 20 = 9960 deformations necessary for this experiment were performed on a computing grid of desktop machines.

Weighting parameters λ_1 and λ_2 were again chosen independently from the target image. This was done using leave-one-out parameter-tuning experiments with the training images. This procedure was comparable to the method described in section 2.3.4, but for these parameter-tuning experiments all 20 training images from set I were available, instead of 19. The set of weights that gave the highest mean SI over all training images was then used to segment all target images.

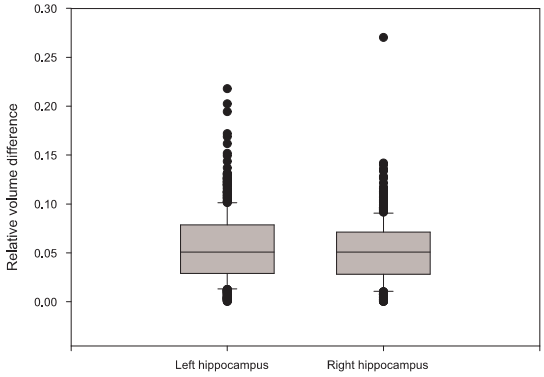


Figure 2.9: *Boxplot showing the relative volume difference between the graph cut and multi-atlas segmentations. Shown are the median values, 25% and 75% quartiles, and the outliers*

Visual rating

The quality of the segmentations was visually inspected by an expert observer (T.d.H.) and rated with a score of 0 (poor), 1 (reasonable), or 2 (good). A rating of 'good' meant no apparent error. The segmentation deviated no more than one or two voxels from the real boundary. A rating of 'reasonable' indicates noticeable, but small errors. This rating is used for a limited number of errors that deviate into the background, for example when the foreground segmentation includes part of the amygdala or the temporal horn of the ventricle. A 'poor' segmentation showed large deviations, for example when it crossed over to a neighboring gyrus. The observer also indicated if the method produced an over- or undersegmentation.

372 (75%) of the 498 left hippocampi were rated as 'good' by the observer, 80 (16%) as 'reasonable', and 46 (9%) as 'poor'. For the right hippocampi these frequencies were 415 (83%), 45 (9%), and 38 (8%). Oversegmentations were far more common than undersegmentations according to the observer: 83 of the left hippocampus segmentations were too large, and 8 too small. On the right side, the observer reported oversegmentation in 44 images and undersegmentation in 4.

The majority of the segmentations rated as 'poor' were cases of oversegmentation of the neighboring gray matter structures, as described in section 2.3.4. The frequency of these errors in image set II was about 8%, which is comparable to their occurrence in set I.

Qualitative comparison with multi-atlas segmentation

A further comparison was made between the proposed graph cut method and the multi-atlas method by visual inspections of all subjects for which the relative volume difference between both methods showed a large disagreement. For all images

in set II, the multi-atlas-based segmentations were computed according to the 'winner takes all' principle. In case of a tie, a voxel was labeled as background. Subsequently, volumes were derived from these results and compared with the volumes measured with the graph-cut-based method using:

$$\left|RV_{gc-ma}\right| = \frac{\left|V(\mathbf{f}_{ma}) - V(\mathbf{f}_{gc})\right|}{V(\mathbf{f}_{gc})}, \tag{2.18}$$

with $V(\mathbf{f}_{gc})$ the hippocampus volume according to the graph cut method, and $V(\mathbf{f}_{ma})$ the volume according to the multi-atlas method.

All segmentation pairs with a $\left|RV_{gc-ma}\right|$ larger than 0.1 were shown for comparison to an expert observer in a blinded fashion. Also included in the experiment was a control group of 40 hippocampi picked at random from the group with a $\left|RV_{gc-ma}\right|$ smaller than 0.1.

The mean relative volume difference $\left|RV_{gc-ma}\right|$ between the two methods was 0.056 for the left hippocampi, and 0.052 for the right. Figure 2.9 shows boxplots of the differences. Of the 498 left hippocampi, 52 (10%) had a $\left|RV_{gc-ma}\right|$ larger than 0.1. For the right hippocampus this occurred in 25 cases (5%).

For the group with large volume differences the observer rated the graph-cut-based method as better in 48(62.3%) of the 77 hippocampi. In 21(27.3%) cases the multi-atlas and graph cut performed equally, and in 8(10.4%) cases the multi-atlas method was rated as best. In the control group, the graph-cut-based segmentation was rated 28 times(70%) as better, 11 times(27.5%) as equally good, and 1 time (2.5%) as worse than the multi-atlas-based segmentation.

Automated versus manual hippocampal volumetry

To assess the validity of the graph-cut-based method, we performed several analyses on the relation between hippocampal volume of image set II and specific determinants and outcomes. Associations were analyzed between the determinants age, sex, and APOE $\varepsilon 4$ and hippocampal volume, as well as between volume and cognitive status, memory performance, and the incidence of dementia after brain MRI. The associations were compared to the associations with manual hippocampal volume measurements of the same data, analyzed in [den Heijer et al., 2002, 2003, 2006].

Global cognitive status was assessed with the Mini-Mental State Examination (MMSE), and memory performance with a verbal memory task [Hackert et al., 2002]. At baseline in 1995-1996 when all 498 persons had brain MRI, none were demented. The subjects in the cohort were repeatedly re-examined (1999-2000, 2001-2002, and 2003-2004). During a total follow-up time of 2891 person-years (mean 5.8 years per person), 48 persons were diagnosed with incident dementia. For details on the dementia screening, see den Heijer et al. [2006]. In short, persons with a MMSE lower than 26 or a Geriatric Mental State Schedule of more than 0 were additionally assessed with the Cambridge examination of mental disorders of the elderly (CAMDEX). Those suspected of dementia based on the CAMDEX were seen

Table 2.2: *Associations between determinants and manually and automatically obtained hippocampal volumes for left and right side.*

	Manual volume*		**Graph cut volume***	
	β-coefficient (95% CI)	p-value	β-coefficient (95% CI)	p-value
Left hippocampus				
Age, per year	-0.012 (-0.017 ; -0.008)	$<$0.001	-0.003 (-0.006 ; 0.000)	0.068
Sex†	0.21 (0.14 ; 0.28)	$<$0.001	0.17 (0.12 ; 0.22)	$<$0.001
APOE ε4 allele‡	-0.10 (-0.19 ; -0.02)	0.02	-0.09 (-0.15 ; -0.03)	0.004
Right hippocampus				
Age, per year	-0.010 (-0.014 ; -0.005)	$<$0.001	-0.005 (-0.008 ; -0.002)	0.003
Sex†	0.22 (0.15 ; 0.29)	$<$0.001	0.04 (-0.01 ; 0.09)	0.13
APOE ε4 allele‡	-0.10 (-0.19 ; -0.01)	0.04	-0.09 (-0.15 ; -0.03)	0.004

Values are regression coefficients (95% CI).
*Volumes are divided by total intracranial volume.
† Positive regression coefficient indicates larger volume in women than in men.
‡ Adjusted regression coefficient for age and sex.

by a neurologist and a neuro-psychologist. For persons not visiting the research center, we used information of the general practitioners and mental health institutions to find potential incident dementia cases. All potential cases were reviewed by a board of a neurologist, a neuro-psychologist, and a research physician.

The manual hippocampal volumes as measured in 1999 were 2.80$\pm$0.43 on the left and 2.86$\pm$0.40 on the right (expressed as ‰of ICV). The graph cut volumes of the same set were 2.71$\pm$0.28 on the left and 2.77$\pm$0.28 on the right (expressed as ‰of ICV). The ICC between the manual and automated volumes was 0.65 for the left and 0.70 for the right. However, as described in section 2.3.1, the manual volumes of set II were obtained from segmentations created with a slightly different segmentation protocol than the training images used for the automated segmentation.

Table 2.2 shows the associations between the determinants and hippocampal volume, both for the manual volume and the graph-cut volumes. Analyses were computed using multiple linear regression, and adjusted for age and sex if needed. The table lists the regression coefficients alongside 95% confidence intervals and p-

Table 2.3: *Associations between both manually and automatically assessed hippocampal volume and outcome variables.*

	MMSE*		**Memory function***		**Incident dementia***	
	β-coefficient (95% CI)	p-value	β-coefficient (95% CI)	p-value	Hazard ratio (95% CI)	p-value
Left hippocampus						
Manual	0.16 (-0.05 ; 0.36)	0.13	0.15 (0.07 ; 0.23)	<0.001	0.39 (0.27 ; 0.57)	<0.001
Graph cut	0.19 (-0.00 ; 0.38)	0.06	0.14 (0.07 ; 0.22)	<0.001	0.45 (0.33 ; 0.61)	<0.001
Right hippocampus						
Manual	0.06 (-0.14 ; 0.26)	0.54	0.13 (0.04 ; 0.21)	0.003	0.38 (0.27 ; 0.54)	<0.001
Graph cut	0.20 (0.02 ; 0.39)	0.03	0.17 (0.10 ; 0.25)	<0.001	0.48 (0.35 ; 0.66)	<0.001

*Values are adjusted regression coefficients (95% CI) or hazard ratios per SD increase in volume (absolute volumes were divided by total intracranial volume). Adjustments were made for age and sex.

values. Older age was associated with smaller hippocampal volumes, both for the manual volumes and the graph cut volumes. However, this relation was weaker for the automated measurements, and not statistically significant on the left side. Women had on average larger volumes, though this association was weaker and not significant for the right graph cut volumes. APOE $\varepsilon 4$ is a well-known genetic risk factor for Alzheimer's disease, and we show here that it is associated with a smaller hippocampal volume, either manually assessed or obtained using the graph cut method.

Table 2.3 lists associations between hippocampal volumes and cognitive measures. We found that a larger hippocampal volume was associated with a higher score on the MMSE and on memory function. This was present for both manual volumes and automated volumes. However, the associations between graph-cut-measured volumes and MMSE were statistically significant, but not between manually measured volume and MMSE scores. Persons with a larger hippocampal volume had a much lower risk to develop dementia. Again, this relation was found both for the manually and automatically measured volumes.

When the segmentations rated as 'poor' in the visual rating experiment were excluded from analysis, the results did not change significantly.

2.4 Discussion and Conclusions

2.4.1 Segmentation quality

The results presented in this work demonstrate that the proposed graph-cut-based method is well-suited for automated hippocampus segmentation. With a mean SI of 0.852 and 0.864 for left and right, the automated segmentations showed good overlap with manual expert segmentations. D_{mean} was below the voxel size with values of 0.4 mm for both sides of the brain. Intraclass correlations between automated and manual volumes exceeded 0.8. Visual rating of the segmentations of 498 unlabeled images also showed excellent results: 83% of the 996 segmented hippocampi were rated as 'good' by a trained observer.

Experiments with the number of training images showed that a mean SI of 0.825 can be obtained with only three training images. With 11 images the SI becomes 0.850. Further increasing the number of training images does little to increase the mean, but does improve robustness by reducing the range.

The regression coefficients and scatter plots indicate that the proposed method suffers from underestimation of large hippocampi compared to the manual segmentation. Visual inspection suggests that these errors are caused by misregistrations of the medial side of the hippocampal head and tail. The proposed method's intensity model can partially correct for these errors, which is supported by the fact that the underestimation is more pronounced in the multi-atlas-based segmentations. A possible explanation for the misregistrations is that large hippocampi tend to have bigger heads and tails with more complicated shapes. Although the set of training images contains a range of hippocampus sizes, the combined result

after registration might be biased to an average shape. The visual rating results seemed to suggest that this problem was less prevalent in image set II: the expert reported 127 oversegmentations versus 12 undersegmentations. However, further inspection showed that undersegmentation did occur in these segmentations, but that their effect were less conspicuous than the cases of oversegmentation.

Some of the oversegmentations were caused by registration errors that venture into neighboring gray matter areas. In these cases the graph-cut-based method produces results inferior to the multi-atlas based method as the the intensity model cannot distinguish between foreground and background. As a consequence it will draw the hippocampus border further away from the real boundary. This type of error occurred in less than 10% of the segmentations of both image sets.

This type of oversegmentation can be detected by comparing the graph-cut-based volume with the result of the thresholded probability map. Images that show a large difference could be automatically detected for visual inspection, and if necessary, corrected. This could be done by manual editing, resorting to the multi-atlas-based segmentation, or removing the misregistered images from the averaged probability map and rerunning the graph cut segmentation.

All segmentations computed in the accuracy experiment also show a small overestimation at the border with the amygdala. This is a consequence of the proposed method's energy model. The boundary with the amygdala is ill-defined on MRI, as this structure has a similar intensity distribution to the hippocampus. The intensity model will therefore consider the voxels of the amygdala as foreground. Furthermore, the regularity model detects no large intensity differences across the boundary and therefore imposes a penalty on placing the edge there. But as the border area is small and the oversegmentation is limited to 20 to 50 voxels, these errors did not have a large effect on the overall accuracy.

This problem could be addressed by separating the background in multiple classes. The intensity model and regularizer could then be adapted to the expected intensities at the hippocampus border. For example, at the interface with the amygdala where the intensity and regularity models carry no additional information, the spatial prior could be given a larger influence on the final segmentation. However, for more than two labels graph cuts are not guaranteed to find a global optimum, although in practice robust solutions are often found [Boykov et al., 2001].

As the graph cut method was evaluated on images acquired with a relatively little-used MRI sequence, the results of the experiments cannot be directly generalized. As the HASTE sequence shows good contrast between gray matter, white matter and cerebrospinal fluid, we do believe that the method will give similar results on scans from other MR sequences with similar contrasts, for example T1-weighted images. However, additional experiments would be necessary to verify this claim.

2.4.2 Application to volumetry

The proposed method showed potential for application in volumetry studies. We found several associations between automatically measured volumes, and poten-

tial determinants and consequences of a small hippocampal volume. These relations included a decreased risk of dementia for subjects with a larger hippocampal volume, positive associations between hippocampal volume and memory function, and a smaller hippocampus volume for APOE $\varepsilon 4$ carriers.

The automatically measured volumes were compared to previously measured manual volumes by computing ICCs and comparing associations. The correlation between manual and graph cut volumes was 0.65 for the left and 0.70 for the right side. However, the ICC might be underestimated, since the segmentation protocol of the training images differed somewhat from the protocol used to obtain the manual volumes. All relations studied showed similar trends for automatically and manually measured volumes, but some differed in statistical significance. The correlations between sex and right hippocampal volume, and age and left hippocampal volume were significant when measured manually, but not when measured with the graph-cut-based method. The smaller effect sizes of the association with age could be explained by the underestimation of relatively large volumes discussed in the previous section. For the associations between left or right volumes and MMSE scores, the manual volumetry study found no significant relations, whereas the automated study did.

Associations with age, sex and MMSE are less established in the literature than hippocampal size and the incidence of dementia. For example, aging has been associated with hippocampal volume loss [Xu et al., 2008], without [Good et al., 2001], or with losses on one side [Greenberg et al., 2008]. Sex difference have been controversial, with some studies showing higher volumes in women [Greenberg et al., 2008], and others not [Good et al., 2001].

To apply the proposed method to an imaging study, specific training data must be available for the sequence and scanner used in that study. The intensity model is based on a supervised classifier, which makes it impossible to use training data from other sequences. Several segmentation methods have been presented in the literature that use an atlas to sample intensity models from the target. A similar strategy could be incorporated in our method to make it more sequence-independent.

Although the proposed method requires no user interaction, its high computational burden can also complicate the application to large image studies. Computing the registrations from 20 different labeled images to a single target image took between 100 and 160 CPU hours. The subsequent computation time of the graph cut was negligible (about 1 second). The majority of computations for this work were therefore done on a grid of desktop computers. In the future, we will investigate reduction of this effort by optimizing the number of registered training images and the control point spacing of the registration.

2.4.3 Comparison to other methods

The graph-cut-based method performed consistently better than a multi-atlas-based approach. The SI and D_{mean} differences between the two methods were statistically significant in paired t-tests. When comparing the accuracy of both methods as a function of the number of training images J, the graph-cut-based method produces

better results for all values of J. This difference decreases from 0.032 for 3 to 0.018 for 19 images used. The direct visual comparison of both methods' volumes on a large image set shows that the the graph cut method gave better or equal results in a majority of cases.

The tendency to underestimate large hippocampal volumes found for the graph-cut-based method is stronger for the multi-atlas volumes. The scatter plots and regression coefficients show that the variance of the manually measured volumes is further reduced when measured with the multi-atlas-based segmentation. Also, the graph-cut-based method's volumetric ICC was much higher.

Several other groups have presented brain segmentation methods with energy models comparable to the model described in this work. However, these techniques rely on local optimizers like iterative conditional modes [Fischl et al., 2002] or mean field approximation [van Leemput et al., 1999] to find a solution. Graph cuts have been used before in brain structure segmentation. However, most of these studies did not provide a quantitative validation [Freedman and Zhang, 2005, Weldeselassie and Hamarneh, 2007]. Song et al. [2006] used graph cuts to iteratively segment brain tissue types and estimate an MR bias field. They validate their results with manual segmentations, but do not apply their method on brain structures.

Table 2.4 compares the method and estimated accuracy of the graph-cut-based method with several alternative, automated hippocampus segmentation methods. The proposed method outperforms atlas-based methods without intensity models like the methods by Heckemann et al. [2006] and by [Gouttard et al., 2007]. On the other hand, the lack of an intensity model makes these methods less dependent on the target image's intensity characteristics. The large difference between the SI reported by Gouttard *et al.* and this work may be explained by the fact that they use images of both normal and autistic children. Moreover, they employ a single registration to align the atlas (created in an unbiased template space) to the target. This has been shown to be less accurate than creating the atlas in the coordinate framework of the target image [Heckemann et al., 2006].

The other studies listed in Table 2.4 all combine information from atlas registration and statistical in a Bayesian framework. Two of these methods produce slightly better SIs than the proposed method [Chupin et al., 2009, Han and Fischl, 2007]. Han and Fischl [2007] use a two-step procedure, which first aligns the atlas, and then computes a complete brain segmentation based on statistical models of intensities and spatial distributions. The most important differences with this work is that their method has a forward model of MR image formation, includes a voxel-based neighborhood model, and provides an explicit model for the structures surrounding the hippocampus. A previous version of this method [Fischl et al., 2002] is used in large hippocampal volumetry studies(for example Goldman et al. [2008]). Chupin et al. [2009] describes a fully-automated, extended version of the semi-automatic technique presented in Chupin et al. [2007]. This method segments the hippocampus and amygdala based on the optimization of an energy function, which includes information from an intensity model, a probability map, a regularizer, and automatically detected anatomical landmarks. It differs from our

graph-cut-based segmentation in the explicit model of the amygdala and the use of landmarks to guide the segmentation.

The SIs reported in Pohl et al. [2007] and Akselrod-Ballin et al. [2007] are lower than the scores reported in this work. The main difference with Pohl et al. [2007] is that it performs a brain segmentation in a hierarchical fashion, with specific weighting of the atlas and intensity terms at each level. Pohl and co-workers used their method to revisit a manual volumetry study, reproducing almost all associations. Akselrod-Ballin et al. [2007] present a graph-based segmentation approach, and use a Bayesian model very similar to our graph-based method. However, their method is based on the aggregation of clusters of increasing size, and uses the model to define a dissimilarity measure. Furthermore, they use affine registration to align a small number of atlases, which might explain the relatively low SI of 0.69.

In summary, we have presented a brain structure segmentation method that combines atlas registration with voxel classification, which gives globally optimal solutions. We have shown that it produces accurate results, and that it outperforms multi-atlas-based segmentation approaches. The method was able to process a large image set without human intervention and shows promise for the analysis of large-scale imaging studies.

Table 2.4: *Hippocampus segmentation results reported in previous work.*

Reference	model*	int. model	atlas		population		accuracy
		scanner-indep.?	N	reg. to template	reg. to target	N	type**	SI±SD [range]
This work	A I R [1]	No	19	N.A.	NR	20	Ag:H	0.85±0.04 [0.76-0.90](L) 0.86±0.02 [0.83-0.89](R)
Atlas-based methods								
Heckemann et al. [2006]	A	Yes	29	N.A.	NR	30	A:H	0.81±0.04 [0.70-0.87](L) 0.83±0.04 [0.71-0.89](R)
Gouttard et al. [2007]	A	Yes	20	NR	NR	20	C:H+P	0.75
Atlas&intensity-based methods								
Fischl et al. [2004]†	A I N	Yes	40	NR	NR	13	A:H	0.87
Akselrod-Ballin et al. [2007]	A I H	No	5	N.A.	Af	18	A:H	0.69
Pohl et al. [2007]	A I H	No	17	Af	NR	50	A:H+P	0.81±0.03(L) 0.81±0.04(R)
Chupin et al. [2009]	A I L R	Yes	15	NR	NR	16	A:H	0.87±0.02 [0.80-0.90]

* A: atlas-based, I: intensity model, N: neighborhood model, R: regularizer, H: hierarchical model, L: landmark-based.
** Ag: aged subjects, C: children, H: healthy subjects, P: psychiatric or autistic subjects.
† These results were reported in Han and Fischl [2007]

References

A. Akselrod-Ballin, M. Galun, J. M. Gomori, A. Brandt, and R. Basri. Prior knowledge driven multiscale segmentation of brain MRI. In *MICCAI 2007, Lecture Notes in Computer Science*, volume 4792, pages 118–126, 2007.

J. Ashburner and K. J. Friston. Unified segmentation. *NeuroImage*, 26(3):839–851, 2005.

Y. Boykov and G. Funka-Lea. Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vis.*, 70(2):109–131, 2006.

Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(9):1124–1137, 2004.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001.

M. Chupin, A. R. Mukuna-Bantumbakulu, D. Hasboun, E. Bardinet, S. Baillet, S. Kinkingnéhun, L. Lemieux, B. Dubois, and L. Garnero. Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: Method and validation on controls and patients with Alzheimer's disease. *NeuroImage*, 34(3):996–1019, 2007.

M. Chupin, A. Hammers, R. S. N. Liu, O. Colliot, J. Burdett, E. Bardinet, J. S. Duncan, L. Garnero, and L. Lemieux. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *NeuroImage*, 46(3):749–761, 2009.

D. Collins, C. Holmes, T. Peters, and A. Evans. Automatic 3-d model-based neuroanatomical segmentation. *Human Brain Mapping*, 3(3):190–208, 1995.

J. G. Csernansky, S. Joshi, L. Wang, J. W. Haller, M. Gado, J. P. Miller, U. Grenander, and M. I. Miller. Hippocampal morphometry in schizophrenia by high dimensional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.*, 95(19):11406–11411, 1998.

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 Trans. Med. Imaging*, 18 (10):909–916, 1999.

T. den Heijer, M. Oudkerk, L. J. Launer, C. M. van Duijn, A. Hofman, and M. M. B. Breteler. Hippocampal, amygdalar, and global brain atrophy in different apolipoprotein e genotypes. *Neurology*, 59(5):746–748, 2002.

T. den Heijer, S. E. Vermeer, R. Clarke, M. Oudkerk, P. J. Koudstaal, A. Hofman, and M. M. B. Breteler. Homocysteine and brain atrophy on MRI of non-demented elderly. *Brain*, 126(Pt 1):170–175, 2003.

T. den Heijer, M. I. Geerlings, F. E. Hoebeek, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler. Use of hippocampal and amygdalar volumes on magnetic resonance imaging to predict dementia in cognitively intact elderly people. *Arch. Gen. Psychiatry*, 63(1):57–62, 2006.

B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

B. Fischl, D. H. Salat, A. J. W. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23 Suppl 1:S69–S84, 2004.

D. Freedman and T. Zhang. Interactive graph cut based segmentation with shape priors. In *CVPR 2005*, pages 755–762, 2005.

A. L. Goldman, L. Pezawas, V. S. Mattay, B. Fischl, B. A. Verchinski, B. Zoltick, D. R. Weinberger, and A. Meyer-Lindenberg. Heritability of brain morphology related to schizophrenia: a large-scale automated magnetic resonance imaging segmentation study. *Biol. Psychiatry*, 63(5):475–483, 2008.

C. D. Good, I. Johnsrude, J. Ashburner, R. N. Henson, K. J. Friston, and R. S. Frackowiak. Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *NeuroImage*, 14(3):685–700, 2001.

S. Gouttard, M. Styner, S. Joshi, R. G. Smith, H. Cody Hazlett, and G. Gerig. Subcortical structure segmentation using probabilistic atlas priors. In *SPIE Medical Imaging 2007: Image Processing*, volume 6512, 2007.

D. L. Greenberg, D. F. Messer, M. E. Payne, J. R. Macfall, J. M. Provenzale, D. C. Steffens, and R. R. Krishnan. Aging, gender, and the elderly adult brain: an examination of analytical strategies. *Neurobiol. Aging*, 29(2):290–302, 2008.

D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, 51(2):271–279, 1989.

V. H. Hackert, T. den Heijer, M. Oudkerk, P. J. Koudstaal, A. Hofman, and M. M. B. Breteler. Hippocampal head size associated with verbal memory performance in nondemented elderly. *NeuroImage*, 17(3):1365–1372, 2002.

J. W. Haller, A. Banerjee, G. E. Christensen, M. Gado, S. Joshi, M. I. Miller, Y. Sheline, M. W. Vannier, and J. G. Csernansky. Three-dimensional hippocampal MR

morphometry with high-dimensional transformation of a neuroanatomic atlas. *Radiology*, 202(2):504–510, 1997.

A. Hammers, R. Heckemann, M. J. Koepp, J. S. Duncan, J. V. Hajnal, D. Rueckert, and P. Aljabar. Automatic detection and quantification of hippocampal atrophy on MRI in temporal lobe epilepsy: a proof-of-principle study. *NeuroImage*, 36(1): 38–47, 2007.

X. Han and B. Fischl. Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Trans. Med. Imaging*, 26(4):479–486, 2007.

R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.

M. A. Ikram, H. A. Vrooman, M. W. Vernooij, F. van der Lijn, A. Hofman, A. van der Lugt, W. J. Niessen, and M. M. B. Breteler. Brain tissue volumes in the general elderly population. the Rotterdam Scan Study. *Neurobiol. Aging*, 29(6):882–890, 2008.

D. V. Iosifescu, M. E. Shenton, S. K. Warfield, R. Kikinis, J. Dengler, F. A. Jolesz, and R. W. McCarley. An automated registration algorithm for measuring MRI subcortical brain structures. *NeuroImage*, 6(1):13–25, 1997.

S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*, 29(1):196–205, 2010.

V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2):147–159, 2004.

S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.

K. O. McGraw and S. P. Wong. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods*, 1(1):30–46, 1996.

K. M. Pohl, J. Fisher, W. E. L. Grimson, R. Kikinis, and W. M. Wells. A Bayesian model for joint segmentation and registration. *NeuroImage*, 31(1):228–239, 2006.

K. M. Pohl, S. Bouix, M. Nakamura, T. Rohlfing, R. W. McCarley, R. Kikinis, W. E. L. Grimson, M. E. Shenton, and W. M. Wells. A hierarchical algorithm for MR brain image parcellation. *IEEE Trans. Med. Imaging*, 26(9):1201–1212, 2007.

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

J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging*, 17 (1):87–97, 1998.

Z. Song, N. J. Tustison, B. B. Avants, and J. C. Gee. Integrated graph cuts for brain MRI segmentation. In *MICCAI 2006, Lecture Notes in Computer Science*, volume 4191, pages 831–838, 2006.

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

B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard. Image registration via level-set motion: applications to atlas-based segmentation. *Med. Image Anal.*, 7(1):1–20, 2003.

Y. Weldeselassie and G. Hamarneh. DT-MRI segmentation using graph cuts pages. In *SPIE Medical Imaging 2007: Image Processing*, volume 6512, 2007.

Y. Xu, D. J. Valentino, A. I. Scher, I. Dinov, L. R. White, P. M. Thompson, L. J. Launer, and A. W. Toga. Age effects on hippocampal structural changes in old men: the haas. *NeuroImage*, 40(3):1003–1015, 2008.

J. Zhou and J. C. Rajapakse. Segmentation of subcortical brain structures using fuzzy templates. *NeuroImage*, 28(4):915–924, 2005.

Chapter 3

Atlas- and appearance-based brain structure segmentation

This chapter is based on:
Fedde van der Lijn, Marleen de Bruijne, Stefan Klein, Tom den Heijer, Yoo Young Hoogendam, Aad van der Lugt, Monique M.B. Breteler and Wiro J. Niessen. "Automated brain structure segmentation based on atlas registration and appearance models". *Submitted*.

Abstract

This chapter describes a novel method for brain structure segmentation in magnetic resonance images that combines information about a structure's location and its appearance. The spatial model is implemented by registering multiple atlas images to the unlabeled target image and creating a spatial probability map. The structure's appearance is modeled by a classifier based on Gaussian scale-space features. These components are combined with a regularization term in a Bayesian framework that is globally optimized using graph cuts. The incorporation of the appearance model enables the method to segment structures with complex intensity distributions and increases its robustness against errors in the spatial model. The method is tested in leave-one-out experiments on two datasets acquired with different magnetic resonance sequences, in which the hippocampus and cerebellum were manually segmented by an expert. Furthermore, the method is compared to three other segmentation techniques based on atlas registration, atlas registration plus an intensity model, and Freesurfer, which were applied to the same data. Results show that the atlas&appearance-based method produces accurate results with mean Dice similarity indices of 0.95 for the cerebellum, and 0.87 for the hippocampus. This was significantly better than the atlas-based method and comparable to the atlas&intensity-based method, but the atlas&appearance-based method is more widely applicable and robust. The results were compared to Freesurfer by computing correlations between the automated and manual volume measurements. These were found to be very similar for both methods.

3.1 Introduction

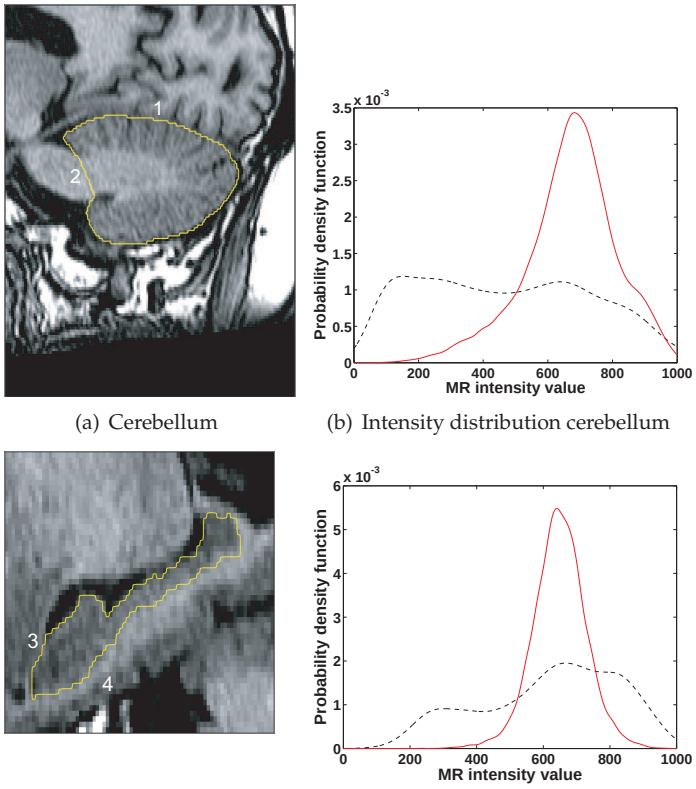
In Chapter 2 an intensity model was used to improve the accuracy and robustness of atlas-based segmentation. Like many other atlas&intensity-based techniques, this method uses a global intensity model to determine whether a voxel belongs to the foreground or background class [Chupin et al., 2009, Fischl et al., 2002, Pohl et al., 2006, Wolz et al., 2010]. However, most brain structures have one or more neighbors with similar intensity characteristics, which results in partially overlapping foreground and background distributions. Background voxels with foreground intensities (or vice versa) will therefore be mislabeled by the intensity model, unless the atlas information is very strong.

As a result, atlas&intensity-based methods are unsuited to segment structure with complex, spatially varying intensity patterns like the cerebellum (Figure 3.1). Large numbers of background voxels at the interface with the cerebrum and the brainstem will be considered as foreground. This limits the applicability of atlas &intensity-based techniques. Structures with relatively simple intensity patterns like the hippocampus show limited overlap, and can therefore usually be accurately segmented. But when applied to the hippocampus, atlas&intensity-based methods remain vulnerable to registration errors that push the atlas into background areas like the enthorinal cortex, amygdala or parahippocampal gyrus (Figure 3.1).

In this chapter we present a novel segmentation method that can overcome these limitations by combining atlas-based segmentation with an appearance model. Previous work has shown the potential of modeling local image appearance of brain structures with high-dimensional feature vectors of Haar filters or Gaussian derivatives at different scales [Arzhaeva et al., 2007, Morra et al., 2008, 2010, Powell et al., 2008, van der Lijn et al., 2009]. The appearance model enables the segmentation of structures with more complex intensity patterns and can potentially increase the method's robustness to registration errors.

The proposed method uses multiple atlas registrations to construct a spatial probability map that models the location of the brain structure in an unlabeled MR image. The appearance of the structure is described by a voxel classifier based on Gaussian scale space features. The smoothness of the result is controlled with a regularization term. The spatial, appearance and regularization terms are then combined in a posterior probability function that can be globally maximized using graph cuts [Boykov et al., 2001, Greig et al., 1989].

The method is evaluated by segmenting the cerebellum and hippocampus in two MRI datasets that were acquired with different scanners and sequences. We determined its accuracy by comparing the results to manual segmentations. The performance of the method was also compared with that of three other techniques based on atlas registration [Heckemann et al., 2006], atlas registration plus an intensity model [van der Lijn et al., 2008], and a combination of atlas registration and a local intensity and neighborhood model [Fischl et al., 2002].



(a) Cerebellum (b) Intensity distribution cerebellum

(c) Hippocampus (d) Intensity distribution hippocampus

Figure 3.1: *Manually segmented T1-weighted images of a cerebellum (sagittal view) and a hippocampus (sagittal view) together with the intensity distributions sampled from the structure (solid curves) and background (dashed curves). The cerebellum distribution overlaps with the intensities of the cerebrum (1) and brainstem (2). In the hippocampus image background voxels from the amygdala (3) and parahippocampal gyrus (4) have similar intensities as the foreground.*

3.2 Method

The segmentation of an unlabeled target image is equivalent to finding the label field $\mathbf{f}$ with the maximum posterior probability given the image information $\mathbf{i}$:

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{i}). \tag{3.1}$$

As we will consider binary segmentations, $\mathbf{f}$ is a vector containing a label $f_m \in \{0,1\}$ for every voxel m in the set $\mathcal{M}$ of voxel locations in the image. Vector $\mathbf{i}$ consists of the intensity values i_m for all voxel locations $\mathcal{M}$.

Explicitly modeling the joint posterior probability $p(\mathbf{f}|\mathbf{i})$ would be feasible only for the smallest of images because of the exponential amount of possible label con-

figurations. However, we can simplify the computation of $p(\mathbf{f}|\mathbf{i})$ by assuming that the label f_m conditioned on the image intensities $\mathbf{i}$ depends only on the labels of its neighbors $n \in \mathcal{N}_m$. This allows us to approximate Equation 3.1 as a Discriminative Random Field (DRF) with one- and two-voxel clique potentials [Kumar and Hebert, 2003, 2006]:

$$p(\mathbf{f}|\mathbf{i}) \approx \frac{1}{Z} \exp\left(\sum_{m \in \mathcal{M}} \lambda_1 A(f_m, \mathbf{i}) + \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}_m} I(f_m, f_n, \mathbf{i}) \right), \tag{3.2}$$

in which we shall assume the terminology of [Kumar and Hebert, 2003, 2006] and call $A(f_m, \mathbf{i})$ the association potential and $I(f_m, f_n, \mathbf{i})$ the interaction potential. λ_1 is a weight parameter that scales the association potential with respect to the interaction potential. Z is a constant that normalizes the summed posterior probabilities of all possible label configurations to 1. As we are only interested in the label configuration $\hat{\mathbf{f}}$ that gives the maximum posterior probability, we can disregard this term in the optimization of Equation 3.2. The association potential is proportional to the log probability that a single voxel assumes a foreground or background label, given all intensity values of the image. This term is based on statistical models of the brain structure's location and appearance. The interaction potential models the relation between two neighboring voxels, given $\mathbf{i}$. In this work, it is implemented as a regularizer that promotes piecewise smooth segmentations.

In the following subsection the association potential is described, which consists of three components: an appearance model, a spatial model, and a global prior term, which will be detailed in separate subsections. Subsequently, the interaction potential is described. Finally, the section is completed with an explanation of the methods used to compute the maximum a posteriori solution and to select the model parameters.

3.2.1 Association potential

In this work, the association potential has the following form:

$$A(f_m, \mathbf{i}) = \ln p_A(f_m|\mathbf{i}), \tag{3.3}$$

in which p_A is the association probability function that can be written as:

$$p_A(f_m|\mathbf{i}) = \frac{p_{gp}(f_m) \cdot p_{model}(f_m|\mathbf{i})}{p_{gp}(f_m = 0) \cdot p_{model}(f_m = 0|\mathbf{i}) + p_{gp}(f_m = 1) \cdot p_{model}(f_m = 1|\mathbf{i})}. \tag{3.4}$$

In this equation $p_{model}(f_m|\mathbf{i})$ is a probability function that contains the spatial and appearance models. $p_{gp}(f_m)$ is a global prior term which affects the probability of foreground voxel labels in the entire image.

The model term $p_{model}(f_m|\mathbf{i})$ is defined as:

$$p_{model}(f_m|\mathbf{i}) = \frac{p_{app}(f_m|\boldsymbol{\xi}_m(\mathbf{i})) \cdot p_s^{\lambda_2}(f_m)}{p_{app}(f_m = 0|\boldsymbol{\xi}_m(\mathbf{i})) \cdot p_s^{\lambda_2}(f_m = 0) + p_{app}(f_m = 1|\boldsymbol{\xi}_m(\mathbf{i})) \cdot p_s^{\lambda_2}(f_m = 1)}, \tag{3.5}$$

in which $p_{app}(f_m|\boldsymbol{\xi}_m(\mathbf{i}))$ is an appearance model probability map, based on an F-dimensional vector $\boldsymbol{\xi}_m(\mathbf{i})$ of appearance features extracted from the image in the neighborhood of m. It describes the probability of label f_m, based on the appearance at location m. $p_s(f_m)$ is a spatial probability map, which records the probability of encountering label f_m at voxel location m according to a spatial model. λ_2 determines the balance between the spatial and appearance models.

3.2.2 Appearance model

The appearance model $p_{app}(f_m|\boldsymbol{\xi}_m(\mathbf{i}))$ is constructed for the unlabeled target image u by applying a k-nearest-neighbor (knn) voxel classifier operating in the F-dimensional feature space. The classifier is trained by extracting foreground and background samples from a set of J manually labeled training images $\mathcal{T} = \{t_1, \ldots, t_J\}$. The probability that a voxel with a feature vector $\boldsymbol{\xi}_m$ has class label f_m is given by:

$$p_{app}^{(u;\mathcal{T})}(f_m|\boldsymbol{\xi}_m(\mathbf{i})) = \frac{k_{f_m}(\boldsymbol{\xi}_m(\mathbf{i})) + 1}{k+2}, \tag{3.6}$$

in which $k_{f_m}(\boldsymbol{\xi}_m(\mathbf{i}))$ counts the number of training samples with label f_m among the k nearest neighbors of the point $\boldsymbol{\xi}_m(\mathbf{i})$ in the F-dimensional feature space. The superscript $(u;\mathcal{T})$ is used throughout this paper whenever it is important to explicitly specify the target image and the training set. Since the knn classifier makes no assumptions about the distribution of $\boldsymbol{\xi}_m(\mathbf{i})$, it can model complex decision boundaries and has been shown to be effective in brain structure segmentation [Arzhaeva et al., 2007].

We used a moderated knn instead of the standard expression $k_{f_m}(\boldsymbol{\xi}_m(\mathbf{i}))/k$ to ensure that no voxel labels could be 'vetoed' by the appearance model [Kittler and Alkoot, 2002]. The foreground samples were obtained by random sampling from all voxels that were labeled as part of the structure of interest. An equal number of background samples was randomly extracted from a band around the manual segmentation. In all experiments k was set to 10, and we used a fast implementation based on approximate nearest neighbor searching (with an error bound ϵ of one) [Arya et al., 1998]).

The appearance was modeled with Gaussian scale-space features. These were a Gaussian-filtered version of the original image, 1^{st}, and 2^{nd} order Gaussian derivatives in all axis directions, gradient magnitude, Laplacian, Gaussian curvature, and the three eigenvalues of the Hessian. Including the original intensity values, we used $1 + 16n_\sigma$ features for classification (with n_σ the number of scales). The features were independently standardized to zero mean and unit variance. Unlike other segmentation methods that use high-dimensional feature vectors (for example [Arzhaeva et al., 2007, Morra et al., 2008]), we did not include location features; this information is introduced in the model by the spatial component.

A subset of the F most relevant image descriptors was found with sequential forward feature selection, followed by sequential backward selection. In this process, features were added until the area under the ROC curve no longer increased.

Then features were iteratively removed until the performance started to deteriorate. The feature selection was trained on one half of the J training images and its performance was estimated on the other half.

The foreground sample fraction, extent of the background sample area, and the feature scales, were chosen differently for the different image sets and brain structures for which the appearance model was constructed. Their values can be found in Section 3.3.3.

3.2.3 Spatial model

The spatial model was constructed by non-rigidly registering the J training images to the unlabeled target image u. The labels of the training images were then deformed and averaged to create a probability map $p_s(f_m)$:

$$p_s^{(u;\mathcal{T})}(f_m) = \frac{1}{J} \sum_{t_j \in \mathcal{T}} a_m^{(u;t_j)}. \tag{3.7}$$

In this equation $a_m^{(u;t_j)} \in \{0,1\}$ represents the atlas labels of training image t_j deformed to the coordinate frame of target image u and interpolated at voxel location m.

All registrations were computed by first finding an affine transformation followed by a non-rigid transformation parameterized by B-splines. The non-rigid registration step was computed in a multi-resolution fashion with increasing B-spline control point resolution. In all cases mutual information was used as similarity measure. The brain-structure specific registration settings are discussed in Section 3.3.3. All registrations were computed using the Elastix software [Klein et al., 2010].

3.2.4 Global prior

As the appearance model is trained with an equal number of samples for foreground and background, the resulting classifier might not accurately reflect the prior class probabilities. Furthermore, the spatial model might exhibit under- or oversegmentation. The global prior term can compensate for these types of errors by increasing or decreasing the posterior probability of a foreground in the entire image:

$$p_{gp}(f_m) = \begin{cases} \alpha & \text{if } f_m = 0 \\ 1-\alpha & \text{if } f_m = 1 \end{cases} \tag{3.8}$$

with parameter α between 0 and 1. If α has a value of 0.5, the association probability p_A will be unaffected. An α larger than 0.5 decreases the probability of a foreground label for all voxel locations, which decreases the volume of the segmentation. A smaller value increases the foreground probability and the resulting volume.

3.2.5 Interaction potential

As the association potential for each voxel is independent of the others, the resulting segmentation could be noisy. To increase the probability of more smooth segmentations the interaction potential was implemented as follows:

$$I(f_m, f_n, \mathbf{i}) = \begin{cases} 0 & \text{if } f_m = f_n \\ -\frac{1}{2} w_{m,n} \cdot B(\Delta\boldsymbol{\xi}_{m,n}(\mathbf{i})) & \text{if } f_m \neq f_n. \end{cases} \tag{3.9}$$

In this expression $w_{m,n}$ is a distance weight equal to $d_{m,n} / \sum_{l \in \mathcal{N}_m} d_{m,l}$, with $d_{m,n}$ the distance between voxel locations m and n. $B(\Delta\boldsymbol{\xi}_{m,n}(\mathbf{i}))$ is the penalty term for assigning different labels to voxels m and n, which is small if the appearance difference between the voxels is high. The appearance difference is modeled by the Euclidean distance $\Delta\boldsymbol{\xi}_{m,n}(\mathbf{i})$ between the F-dimensional feature vectors $\boldsymbol{\xi}_m(\mathbf{i})$ and $\boldsymbol{\xi}_n(\mathbf{i})$. The penalty is given by a logistic function:

$$B(\Delta\boldsymbol{\xi}_{m,n}(\mathbf{i})) = \frac{1}{1 + \exp(\beta_0 + \beta_1 \Delta\boldsymbol{\xi}_{m,n}(\mathbf{i}))}, \tag{3.10}$$

in which β_0 and β_1 control the offset and slope of the term.

Equation 3.9 promotes smoother segmentations by decreasing the posterior probability of a label configuration in which neighboring voxels have different labels. However, if the feature space distance between neighboring voxels is large, we assume that they belong to different structures. In that case, the logistic function limits the reduction of the posterior probability. This model is a multi-feature version of the gradient-modulated Ising model commonly used in graph cut segmentation methods [Boykov and Funka-Lea, 2006, van der Lijn et al., 2008].

3.2.6 Optimization and parameter learning

The posterior probability function $p(\mathbf{f}|\mathbf{i})$ described above is completely defined by the voxel classifier result p_{app} that models the appearance, the atlas registration result p_s that models the spatial probability, and a parameter vector $\boldsymbol{\theta}$. The latter holds the model's five free parameters: the association weight λ_1, the spatial model weight λ_2, the foreground threshold α, and the parameters of the logistic interaction model β_0 and β_1.

The optimal value of these parameters $\hat{\boldsymbol{\theta}}$ depends heavily on the quality of the appearance and spatial models, which is not known. However, the model quality can be estimated with cross-validation experiments using the manual segmentations of the training images. In this work specifically, we chose $\hat{\boldsymbol{\theta}}$ from a pre-defined range of values $\boldsymbol{\Theta}$ using exhaustive search, which is explained in more detail in Section 3.3.2.

Once the parameters have been chosen, a maximum a posteriori (MAP) solution $\hat{\mathbf{f}}$ can be found by converting Equation 3.1 to an equivalent energy function by taking the negative logarithm. As shown in [Kolmogorov and Zabih, 2004] this function can be globally minimized using graph cuts [Boykov et al., 2001]. In this

work we used the Maxflow algorithm to compute the graph cuts [Boykov and Kolmogorov, 2004].

3.3 Experiments and Results

The method was tested by segmenting the cerebellum and the hippocampus in T1-weighted images. The cerebellum exhibits a complex, spatially varying intensity pattern, whereas the hippocampus has a simpler uniform intensity distribution. To assess the method's ability to handle different MR sequences we also segmented the hippocampus in Half-Fourier Acquisition Single-Shot Turbo Spin Echo (HASTE) images with a lower resolution. The segmentation accuracy was determined by computing overlap and distance measures with respect to manual segmentations in a leave-one-out experiment. The atlas&appearance-based method was also compared to three alternative techniques. The subjects and image data are described in more detail in 3.3.1. The parameter learning procedure and data-specific implementation details are given in Section 3.3.2. The experiments are detailed in Section 3.3.3. Finally, the results are described in Section 3.3.4.

3.3.1 Image data

We used two MR datasets from the Rotterdam Scan Study. The subjects were taken from two different study cohorts and selected to cover the cohorts' variability in age, sex, and global brain size (as measured with an automated brain tissue segmentation method).

Set I consisted of 10 women and 8 men with a mean age of 74.2$\pm$7.9 years. These images were made with a 1.5T GE scanner. We used a 3D T1-weighted sequence (inversion time 400 ms, repetition time 14.8 ms, time to echo 2.8 ms, 96 axial slices of 1.6 mm interpolated to 192 slices of 0.8 mm, acquisition matrix 416x256, field of view 250x250 mm). The final voxel size was $0.49 \times 0.49 \times 0.8$ mm. The hippocampi and cerebellum in set I were segmented by one observer (Y.Y.H.) under supervision of a neurologist (T.d.H.) and a neuro-radiologist (A.v.d.L.). These structures were delineated every other slice (in sagittal view for the cerebellum and in coronal view for the hippocampus). Linear interpolation was used to obtain segmentations for the skipped slices. We shall refer to the hippocampus segmentations of this set as I-HC and to the cerebellum segmentations as I-CRBL.

The low-resolution set II consisted of 9 women and 11 men with a mean age of 74.6$\pm$8.2 years. These images were acquired on a 1.5T Siemens scanner with a custom-made 3D HASTE sequence (inversion time 440 ms, repetition time 2800 ms, 128 contiguous sagittal slices of 1.25 mm, acquisition matrix 192x256, field of view 256x256 mm). Two HASTE modules were sequentially acquired after the inversion pulse (effective echo time of 29 ms and 440 ms) of which the first was used in this work. The final voxel size of these images was $1.25 \times 1.0 \times 1.0$ mm. In these images the hippocampi were delineated on coronal slices by two raters. Fifteen images were segmented by an expert neurologist (T.d.H.) and five by a trained observer

(Y.Y.H.) under supervision of a neurologist (T.d.H). We shall refer to these images and their labels as the II-HC set. The images from both sets were corrected for non-uniformities using N3 [Sled et al., 1998].

3.3.2 Segmentation procedure

The atlas&appearance-based method was applied to the I-CRBL, I-HC, and the II-HC sets. These segmentations were performed in a leave-one-out experiment consisting of three steps. First for every image $t_j \in \mathcal{T}$ an appearance model $p_{app}^{(t_j;\mathcal{T}_j)}$ and a spatial model $p_s^{(t_j;\mathcal{T}_j)}$ was created using the remaining subjects' scans $\mathcal{T}_j = \mathcal{T} \setminus \{t_j\}$ as training images. Secondly, based on these models, segmentations $\hat{\mathbf{f}}^{(t_j)}(\boldsymbol{\theta})$ were computed for all parameter values $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all target images $t_j \in \mathcal{T}$. We then measured the Dice similarity indices $DSI(\hat{\mathbf{f}}^{(t_j)}(\boldsymbol{\theta}), \mathbf{g}^{(t_j)})$ between $\hat{\mathbf{f}}^{(t_j)}(\boldsymbol{\theta})$ and the manual segmentations $\mathbf{g}^{(t_j)}$. This records the segmentation accuracy as a function of the parameters $\boldsymbol{\theta}$ and target image t_j. In the third step, the optimal parameters $\hat{\boldsymbol{\theta}}^{(t_j)}$ were selected for target t_j by finding the parameters that gave the highest mean similarity index computed over all other images $\mathcal{T}_j$. With these parameters the segmentation $\hat{\mathbf{f}}^{(t_j)}(\hat{\boldsymbol{\theta}}^{(t_j)})$ was computed. In this way, parameter learning for the segmentation of t_j was never based on spatial or appearance models constructed in the coordinate system of t_j. The whole procedure is summarized in figure 3.2.

To create the appearance model for I-CRBL, 1% of the manually labeled foreground voxels in the training images were sampled. The background samples were taken from a band up to 10 mm around the foreground. We used $n_\sigma = 4$ with equal logarithmic intervals between 0.5 and 5 mm. The I-HC appearance model was based on 5% of the foreground voxels and a background band of 4 mm. Five scales were used between 0.5 and 5 mm. The sampling parameters of II-HC were identical to those of I-HC, but because of the lower resolution of these images we used three scales between 1 and 5 mm.

The spatial models for I-CRBL and II-HC were based on registrations driven by mutual information computed over the entire image. For I-HC the registration was initialized with the deformation field computed for the I-CRBL set, and further refined in a region of interest around the hippocampus. The registration settings can be found in the Elastix parameter database [1].

The interaction potential was based on a 26-voxel neighborhood for the hippocampi. To reduce computation cost and memory requirement of the graph cut we used a 6-voxel neighborhood for the cerebellum. For the same reason, separate subimages were created around the left and right hippocampus and the cerebellum after construction of the spatial and appearance models. These cropped images were based on bounding boxes around the thresholded spatial probability maps of the structures.

[1] http://elastix.isi.uu.nl/wiki

Figure 3.2: *The parameter learning procedure. See text for more details.*

for $t_j \in \mathcal{T}$ **do**
    Construct appearance model $p_{app}^{(t_j;\mathcal{T}_j)}$
    Construct spatial model $p_s^{(t_j;\mathcal{T}_j)}$
end for
for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ **do**
    for $t_j \in \mathcal{T}$ **do**
        Compute posterior probability function $p^{(t_j;\mathcal{T}_j)}(\boldsymbol{\theta})$ based on $p_{app}^{(t_j;\mathcal{T}_j)}$, $p_s^{(t_j;\mathcal{T}_j)}$, and $\boldsymbol{\theta}$
        Compute MAP label configuration $\hat{\mathbf{f}}^{(t_j)}(\boldsymbol{\theta})$
        Compute $DSI(\hat{\mathbf{f}}^{(t_j)}(\boldsymbol{\theta}), \mathbf{g}^{(t_j)})$
    end for
end for
for $t_j \in \mathcal{T}$ **do**
    Compute $\overline{DSI}^{(t_j)}(\boldsymbol{\theta}) = 1/N \sum_{t_k \in \mathcal{T}_j} DSI(\hat{\mathbf{f}}^{(t_k)}(\boldsymbol{\theta}), \mathbf{g}^{(t_k)})$
    Find $\hat{\boldsymbol{\theta}}^{(t_j)} = \arg\max_{\boldsymbol{\theta}} \overline{DSI}^{(t_j)}(\boldsymbol{\theta})$
    Compute posterior function $p^{(t_j;\mathcal{T}_j)}(\hat{\boldsymbol{\theta}}^{(t_j)})$ based on $p_{app}^{(t_j;\mathcal{T}_j)}$, $p_s^{(t_j;\mathcal{T}_j)}$, and $\hat{\boldsymbol{\theta}}^{(t_j)}$
    Compute MAP label configuration $\hat{\mathbf{f}}^{(t_j)}(\hat{\boldsymbol{\theta}}^{(t_j)})$
end for

Computation time of a registration of one atlas image to the target image was approximately ten CPU minutes on the node of a 64-bit Linux cluster. As a result the construction of the spatial model took three CPU hours for I-CRBL, six CPU hours for I-HC, and 3.5 CPU hours for II-HC. The atlas registrations were performed in parallel to reduce computation time. The computation of the appearance model took approximately 0.5 CPU hour per image of I-HC, 1 CPU hour per image of I-CRBL, and 2 CPU minutes per image of II-HC. Constructing and maximizing the posterior probability function was done within a second for the hippocampi and in two minutes for the cerebellum on a desktop computer. The parameter learning took about four days for I-CRBL, two days for I-HC, and one day for II-HC.

3.3.3 Experiments

The atlas&appearance-based method was validated by comparing the results of the leave-one-out segmentations to the manual labelings. We used the following volumetric quality measures: the Dice similarity index (DSI), defined as:

$$DSI = \frac{2V(\mathbf{f} \cap \mathbf{g})}{V(\mathbf{f}) + V(\mathbf{g})}, \tag{3.11}$$

the Jacquard similarity index (JSI), given by:

$$JSI = \frac{V(\mathbf{f} \cap \mathbf{g})}{V(\mathbf{f} \cup \mathbf{g})}, \tag{3.12}$$

the relative volume difference (RV), defined by:

$$RV = \frac{V(\mathbf{f}) - V(\mathbf{g})}{V(\mathbf{g})}, \tag{3.13}$$

and the volumetric, two-way random, absolute agreement, intraclass correlation coefficient (ICC) [McGraw and Wong, 1996] between $V(\mathbf{g})$ and $V(\mathbf{f})$. In these expressions, $V(\mathbf{f})$ and $V(\mathbf{g})$ are the volumes of the automated segmentation $\mathbf{f}$ and the manual segmentation $\mathbf{g}$.

We also computed two surface-based measures: the maximum and mean surface distance D_{max} and D_{mean}. The maximum distance is given by:

$$D_{max} = \max\left\{\delta(\mathbf{f}, \mathbf{g}), \delta(\mathbf{g}, \mathbf{f})\right\}, \tag{3.14}$$

with $\delta(\mathbf{f}, \mathbf{g})$ a set that contains the distances between every surface voxel in automated segmentation $\mathbf{f}$, and the closest surface voxel in the manual segmentation $\mathbf{g}$. The mean surface distance is defined by:

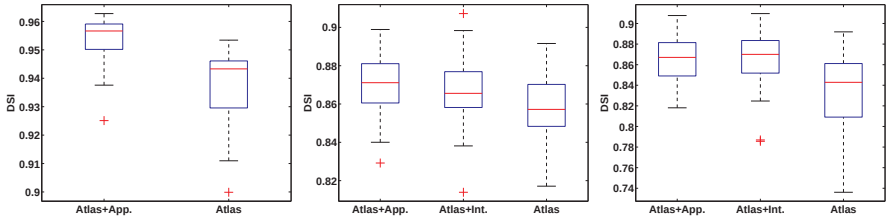
$$D_{mean} = \frac{\bar{\delta}(\mathbf{f}, \mathbf{g}) + \bar{\delta}(\mathbf{g}, \mathbf{f})}{2}, \tag{3.15}$$

with $\bar{\delta}(\mathbf{f}, \mathbf{g})$ the mean of set $\delta(\mathbf{f}, \mathbf{g})$ computed over all surface voxels of $\mathbf{f}$.

To ascertain whether the multi-feature appearance model improves results compared to a model based on MR intensities only, we also segmented the I-HC and II-HC images with the atlas&intensity-based method published in van der Lijn et al. [2008]. This method combines a spatial model, an MR intensity model, and a regularizer in an energy framework that is optimized by graph cuts. I-CRBL was not segmented because the intensity model cannot adequately separate the structure's foreground and background intensities. The results were compared with the manual segmentations using the quality measures listed above.

The atlas&intensity-based segmentations were obtained using the same spatial model as the atlas&appearance results. The intensity model for the unlabeled target image t_j was based on a Parzen classifier trained on intensity values extracted from the manually labeled images $\mathcal{T}_j$. Finally, we used an identical regularizer as in van der Lijn et al. [2008]. The model described in van der Lijn et al. [2008] did not include a global prior p_{gp}, so we added a similar term to the atlas&intensity-based method. The resulting model had three free parameters (equivalent to the λ_1, λ_2, and α), which were optimized in the same way as described in Section 3.3.2.

To assess the added value of the appearance model and the interaction potential, we also compared the performance of the proposed method to that of a multi-atlas-based segmentation [Heckemann et al., 2006]. This method was applied to all three datasets and validated using the same quality measures.



(a) I-CRBL (b) I-HC (c) II-HC

Figure 3.3: *Boxplots of the different methods' DSI scores measured in the three validation sets.*

The atlas-based segmentations were computed by thresholding the spatial model $p_s^{(t_j;\mathcal{T}_j)}$ at value α. This threshold value was chosen based on the training data using a similar procedure as described in Section 3.3.2. We chose to select the threshold based on the training data instead of using a fixed value of 0.5, to make the results better comparable to the proposed method with its global prior p_{gp} term.

All scores are reported as mean $\pm$ standard deviation [min;max]. These statistics were computed over the left- and right-side structures of all images, so N was 36 for the I-HC and I-CRBL sets, and 40 for the II-HC set. We used Kruskal-Wallis signed rank tests to ascertain whether the atlas&appearance-, atlas&intensity-, and atlas-based methods had equal median scores. Additionally, the volume estimates were evaluated by plotting the automatically measured volumes against the manual volumes, and fitting a linear model through this data using linear regression.

Finally, the atlas&appearance-based method was also compared to Freesurfer [Fischl et al., 2002]. Freesurfer is able to segment a large range of brain structures by supplementing spatial information with location-specific intensity and neighborhood models. As this method uses its own definitions of hippocampus and cerebellum, a direct comparison with manual segmentations defined according to a different protocol would not be very informative. We therefore compared the two-way random, consistency ICC between the manually measured volumes of I-HC and I-CRBL and the volumes measured with the automated methods. As Freesurfer can only handle T1-weighted images, we applied this method to I-CRBL and I-HC. The default settings were used.

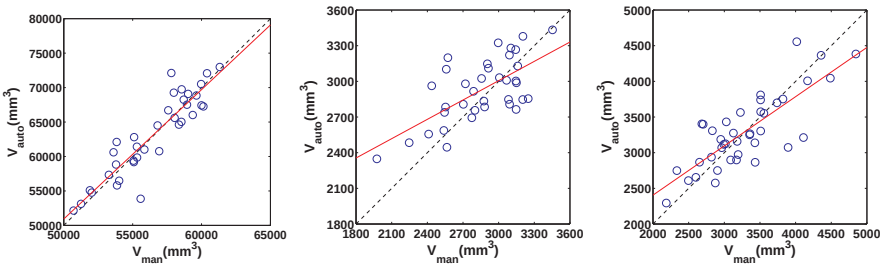
3.3.4 Results

Table 3.1 shows the quality scores of the atlas&appearance, atlas&intensity, and the atlas models for all three validation sets. Figure 3.3 compares the different methods' DSI scores per dataset. When looking at the I-CRBL, the atlas&appearance-based method yielded higher scores across the board, compared to using the atlas alone. The performance of all three methods is very high on the I-HC set, although the atlas-based method never produces the best result. In the II-HC set the atlas&intensity and atlas&appearance methods both perform better than the atlas-based method, but the difference between these two methods is small.

Table 3.1: *Evaluation measures for the atlas&appearance-, atlas&intensity-, and atlas-based methods.*

	atlas&appearance	**atlas&intensity**	**atlas**	**KW-test**
I-CRBL				
DSI	0.954±0.008 [0.925 ; 0.963]		0.937±0.013 [0.900 ; 0.953]	$p < 0.001$
JSI	0.911±0.015 [0.861 ; 0.928]		0.882±0.023 [0.818 ; 0.911]	$p < 0.001$
RV	0.003±0.039 [-0.119 ; 0.099]		0.007±0.057 [-0.145 ; 0.121]	$p = 0.55$
ICC [95%CI]	0.912 [0.853 ; 0.954]		0.817 [0.671 ; 0.902]	-
D_{mean}(mm)	0.50±0.10 [0.38 ; 0.79]		0.68±0.15 [0.50 ; 1.09]	$p < 0.001$
D_{max}(mm)	7.35±3.10 [3.28 ; 16.91]		7.35±1.92 [3.90 ; 10.74]	$p = 0.68$
I-HC				
DSI	0.870±0.017 [0.829 ; 0.899]	0.867±0.018 [0.814 ; 0.907]	0.858±0.017 [0.817 ; 0.892]	$p = 0.008$
JSI	0.771±0.026 [0.708 ; 0.816]	0.766±0.028 [0.686 ; 0.830]	0.752±0.027 [0.691 ; 0.804]	$p = 0.008$
RV	0.031±0.092 [-0.122 ; 0.244]	0.016±0.096 [-0.133 ; 0.273]	0.000±0.079 [-0.156 ; 0.194]	$p = 0.34$
ICC [95%CI]	0.633 [0.391 ; 0.793]	0.609 [0.354 ; 0.779]	0.724 [0.522 ; 0.849]	-
D_{mean}(mm)	0.34±0.06 [0.23 ; 0.53]	0.35±0.64 [0.21 ; 0.58]	0.36±0.06 [0.24 ; 0.52]	$p = 0.35$
D_{max}(mm)	3.69±0.99 [1.93 ; 5.45]	3.86±0.97 [2.33 ; 6.33]	3.53±0.93 [2.18 ; 5.52]	$p = 0.34$
II-HC				
DSI	0.865±0.022 [0.818 ; 0.908]	0.864±0.028 [0.786 ; 0.910]	0.835±0.035 [0.736 ; 0.892]	$p < 0.001$
JSI	0.762±0.034 [0.692 ; 0.831]	0.761±0.043 [0.647 ; 0.834]	0.718±0.051 [0.582 ; 0.805]	$p < 0.001$
RV	0.011±0.109 [-0.218 ; 0.268]	0.015±0.116 [-0.226 ; 0.255]	0.028±0.167 [-0.209 ; 0.412]	$p = 0.54$
ICC [95%CI]	0.797 [0.647 ; 0.887]	0.733 [0.548 ; 0.850]	0.485 [0.205 ; 0.691]	-
D_{mean}(mm)	0.38±0.08 [0.27 ; 0.62]	0.38±0.09 [0.25 ; 0.69]	0.46±0.11 [0.27 ; 0.77]	$p < 0.001$
D_{max}(mm)	4.89±1.77 [2.56 ; 9.39]	5.02±1.63 [2.36 ; 9.01]	4.80±1.61 [2.50 ; 8.95]	$p = 0.79$

Listed are the mean, standard deviation, and range. For the ICC the 95% confidence interval. is given. The p-values were computed using a Kruskal-Wallis signed rank test which tests the hypothesis that all three median scores are equal.



(a) I-CRBL (b) I-HC (c) II-HC

Figure 3.4: *Scatterplots of the volumes measured with the atlas&appearance-based method versus the manually measured volumes for the three validation sets. The regression lines are shown in solid and the perfect segmentation as a dotted line.*

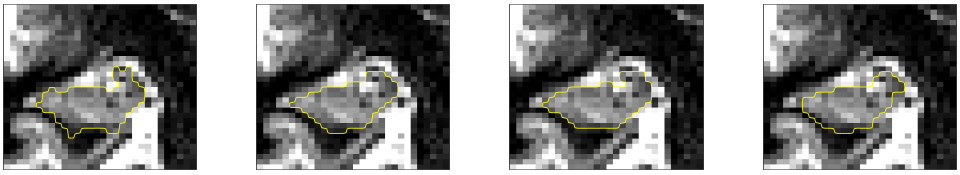
Table 3.2: *Coefficients of the linear models that map the manual volume to volume measured with the atlas&appearance-, atlas&intensity-, and atlas-based methods.*

	atlas&appearance	**atlas&intensity**	**atlas**
I-CRBL			
Intercept, ml	4.0 (-5.4-13.4)		9.0 (-4.4-22.5)
Slope	0.94 (0.79-1.09)		0.86 (0.65-1.08)
I-HC			
Intercept, ml	1.4 (0.8-2.0)	1.4 (0.8-2.1)	1.1 (0.6-1.7)
Slope	0.54 (0.32-0.76)	0.51 (0.28-0.74)	0.60 (0.41-0.79)
II-HC			
Intercept, ml	1.0 (0.5-1.6)	1.5 (1.0-2.0)	2.2 (1.7-2.8)
Slope	0.69 (0.52-0.86)	0.51 (0.40-0.69)	0.32 (0.16-0.49)

Listed are regression coefficients and 95% confidence intervals.

Figure 3.4 shows scatter plots of the volume of the manual segmentation versus that of the atlas&appearance-based segmentation. The regression coefficients of the linear model fitted on the measurements by all three methods are shown in Table 3.2. The atlas&appearance model generally shows the steepest slopes, but all the automated methods have the tendency to underestimate large, and overestimate small volumes for the I-HC and II-HC datasets.

The volumes derived from the Freesurfer segmentations of I-HC and I-CRBL showed an ICC of 0.519 (95%CI 0.233-0.722) and 0.932 (95%CI 0.871-0.965) with respect to the manually measured volumes. The proposed atlas&appearance-based method performed comparably: 0.644 (95%CI 0.404-0.801) for the hippocampus and 0.910 (95%CI 0.832-0.953) for the cerebellum. Note that these were consistency-based ICCs, which explains the small difference with the absolute agreement ICC



(a) manual (b) atlas& appearance (c) atlas&intensity (d) atlas

Figure 3.5: *A coronal slice through segmentations of the same subject from the II-HC set.*

scores listed in Table 3.1.

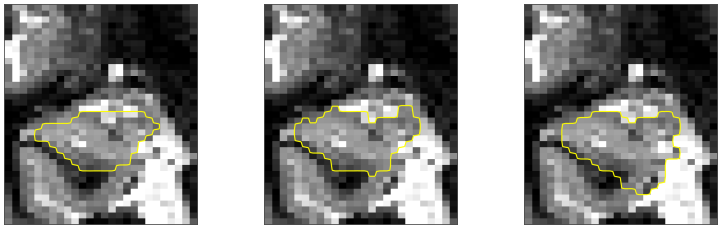
Visual comparison of the different methods showed that the addition of an intensity- or appearance-based component corrects small under- and oversegmentations caused by registration errors (see Figure 3.5). But the atlas&intensity-based method has difficulties dealing with cases where the registration crosses over to neighboring gray matter regions. As can be seen in Figure 3.6(c), the intensity model mistakes these areas for parts of the foreground, which further deteriorates the results. The appearance model on the other hand recognizes that these areas are background and corrected most of these errors (Figure 3.6(b)). This accounts for the removal of the two outliers shown in Figure 3.3(c).

However, in the absence of large registration errors the intensity and appearance components are very comparable. This is especially apparent in set II-HC, where in a small majority of cases the atlas&intensity model outperforms the atlas&appearance model (Figure 3.3(c)). The significant p-value of the Kruskal-Wallis test of the DSI scores listed for this validation set is purely due to the differences between the atlas-based method and the other methods. Unsurprisingly, a post-hoc Wilcoxon signed rank test shows no significant difference between the DSI scores of the atlas&intensity- and atlas&appearance-based methods.

The results of I-CRBL showed some cases of oversegmentation at the posterior border with the skull, caused by registration errors in this area. The proposed method is unable to compensate for these errors as high-intensity voxels in the fatty parts of the skull were considered to be foreground by the appearance model. An extreme example can be seen in Figure 3.7.

3.4 Discussion and Conclusions

The work presented in this paper demonstrates that atlas&appearance-based models can produce robust and accurate segmentations of brain structures with both simple and complex intensity distributions. The proposed method can can handle structures as different in shape and appearance as the hippocampus and the cerebellum. This increases the technique's potential for application to large-scale brain MRI studies compared to atlas&intensity-based methods like [Chupin et al., 2009, Pohl et al., 2006, van der Lijn et al., 2008].



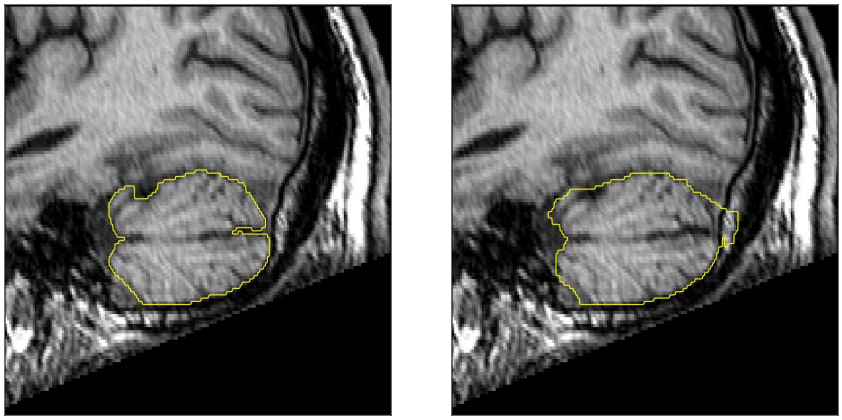
(a) manual (b) atlas& appearance (c) atlas&intensity

Figure 3.6: *Coronal slices through an image of the II-HC set that showed a large oversegmentation by the atlas&intensity-based model. The atlas&appearance-based model avoids these errors.*

The overall segmentation accuracy of the atlas&appearance-based method with respect to manual labelings is good. The DSI and JSI scores on the three validation sets exceeded 0.85 and 0.75. The mean distance error was of the order of the voxel size or smaller. This is comparable to the results of alternative methods reported in the literature [Chupin et al., 2009, Han and Fischl, 2007, Heckemann et al., 2006, Morra et al., 2008, 2010, Pohl et al., 2007, Powell et al., 2008, Wolz et al., 2010]. The atlas&appearance- and atlas-based segmentations did show some large deviations from the manual labellings, especially in the I-CRBL set. This was caused by inclusion of large fissures which were left out of the manual segmentation, while smaller fissures were included by the observer. As this is an error that is very particular to the cerebellum, we expect improved D_{max} scores when the method is applied to other structures.

The volume estimates derived from the proposed segmentation method showed little to no bias. The standard deviation of the volume measurements were 4% and 10% for the cerebellum and hippocampus segmentations respectively. The scatter plot of the automated and manual cerebellar volumes showed no distinct volume-dependent biases. However, for the hippocampal volume measurements there is a tendency to underestimate large, and overestimate small volumes. This bias is likely to be caused by the multi-atlas registration: the atlas-based segmentation without appearance model has a stronger bias towards an average volume. The volumetric ICC was 0.912 for the cerebellum and 0.633 for the hippocampus in set I. The II-HC showed a higher ICC of 0.797, although its RV estimates were quite comparable to that of the I-HC set. This improvement might partly be explained by the larger range of hippocampal volumes in set II. The volumes estimated by the atlas&appearance-based method show a similar correlation to the manual data as the volumes estimated by Freesurfer.

The atlas&appearance-based also showed increased robustness to large registration errors. The appearance model can correct registration errors of the hippocampus when they cross over to gray matter areas like the enthorinal cortex or parahippocampal gyrus. The atlas&intensity-based model cannot distinguish these regions from foreground. Large misregistrations occurred in about 90% of the cases in both



(a) manual (b) atlas&appearance

Figure 3.7: *Example sagittal slice of an oversegmentation in the I-CRBL set caused by the spatial model that is not compensated by the appearance model.*

I-HC and II-HC. This is comparable to the error rate found in an experiment on the entire cohort from which the II-HC set was taken [van der Lijn et al., 2008]. Inclusion of an appearance model would decrease this error rate.

In images where the atlas-based registration did not show any large errors, the proposed method performed better than the strictly atlas-based method as it can correct small registration errors. However, in these cases the atlas&intensity-based method gave comparable results. As long as the spatial model does not venture into the gray matter outside the hippocampus, our experiments suggest that intensity information alone is sufficient to improve the results. Since 90% of the cases had an accurate spatial model, the increased robustness of the appearance model had limited impact on the mean performance scores. For I-HC the spatial model was of such high quality that the atlas-based method performs almost comparably to the methods with additional components.

The atlas&appearance-based method has been successfully applied to two different MR sequences, but it needs sequence-specific training data to do so. This requirement is imposed by the appearance model which needs training data for the classification and feature selection. Since registration with mutual information is relatively robust to intensity differences between atlas and target image, some methods have used the spatial model to sample the intensity model from the target image [Wolz et al., 2010]. Whether this type of approach could be applied to train a high-dimensional appearance model is the subject of further study. Another possibility is to apply an intensity normalization procedure as is done in Freesurfer [Fischl et al., 2004, Han and Fischl, 2007].

The computational costs of the proposed method are high. However, the atlas registrations required for the spatial model can be parallelized, and can be used

to segment multiple structures. Moreover, the extra effort results in a large gain in accuracy and robustness compared to a single atlas registration [Heckemann et al., 2006]. The knn classifier is relatively expensive to apply to a target image, especially compared to classifiers like AdaBoost which are very fast once they have been trained [Morra et al., 2008]. We chose this method because it is flexible and easy to implement, but it could be substituted for a faster classifier, as long as it produces a probabilistic output.

As we had strong spatial and appearance models, we decided to include a relatively simple interaction potential. As shown in Kumar and Hebert [2003, 2006], the DRF framework can easily be extended to incorporate a more complex interaction potential based on a logistic classifier that gives the different features individual weights. In essence, this is a classifier in its own right that labels voxel combinations instead of individual voxels. On the other hand, this model has more parameters which would require a more complex parameter learning strategy.

In conclusion, we have presented a brain structure segmentation method based on atlas registration and multi-feature classification. Because of the classifier's ability to model appearance it can segment structures with both complex and simple intensity distributions. Its accuracy with respect to manual segmentations is good, and comparable or better than existing segmentation methods. Furthermore, the appearance component makes the method more robust to large misregistration compared to atlas&intensity-based methods.

References

S. Arya, D. Mount, N. Netanyahu, R. Silverman, and A. Wu. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of the ACM*, 45(6):891–923, 1998.

Y. Arzhaeva, E. van Rikxoort, and B. van Ginneken. Automated segmentation of caudate nucleus in MR brain images with voxel classification. In T. Heimann, M. Styner, and B. van Ginneken, editors, *3D Segmentation In The Clinic: A Grand Challenge*, pages 65–72, 2007.

Y. Boykov and G. Funka-Lea. Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vis.*, 70(2):109–131, 2006.

Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(9):1124–1137, 2004.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001.

M. Chupin, A. Hammers, R. S. N. Liu, O. Colliot, J. Burdett, E. Bardinet, J. S. Duncan, L. Garnero, and L. Lemieux. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *NeuroImage*, 46(3):749–761, 2009.

B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

B. Fischl, D. H. Salat, A. J. W. van der Kouwe, N. Makris, F. Segonne, B. T. Quinn, and A. M. Dale. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23 Suppl 1:S69–S84, 2004.

D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, 51(2):271–279, 1989.

X. Han and B. Fischl. Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Trans. Med. Imaging*, 26(4):479–486, 2007.

R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.

J. Kittler and F. M. Alkoot. Moderating k-NN classifiers. *Pattern Anal. Appl.*, 5(3): 326–332, 2002.

S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*, 29(1):196–205, 2010.

V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2):147–159, 2004.

S. Kumar and M. Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Proceedings of ICCV*, pages 1150–1159, 2003.

S. Kumar and M. Hebert. Discriminative random fields. *Int. J. Comput. Vision*, 68 (2):179–201, 2006.

K. O. McGraw and S. P. Wong. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods*, 1(1):30–46, 1996.

J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, X. Hua, A. W. Toga, C. R. Jack, M. W. Weiner, P. M. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *NeuroImage*, 43(1):59–68, 2008.

J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson. Comparison of Adaboost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. *IEEE Trans Med Imaging*, 29(1):30–43, 2010.

K. M. Pohl, J. Fisher, W. E. L. Grimson, R. Kikinis, and W. M. Wells. A Bayesian model for joint segmentation and registration. *NeuroImage*, 31(1):228–239, 2006.

K. M. Pohl, S. Bouix, M. Nakamura, T. Rohlfing, R. W. McCarley, R. Kikinis, W. E. L. Grimson, M. E. Shenton, and W. M. Wells. A hierarchical algorithm for MR brain image parcellation. *IEEE Trans. Med. Imaging*, 26(9):1201–1212, 2007.

S. Powell, V. A. Magnotta, H. Johnson, V. K. Jammalamadaka, R. Pierson, and N. C. Andreasen. Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. *NeuroImage*, 39(1):238–247, 2008.

J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging*, 17 (1):87–97, 1998.

F. van der Lijn, T. den Heijer, M. M. B. Breteler, and W. J. Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage*, 43(4):708–720, 2008.

F. van der Lijn, M. de Bruijne, Y. Y. Hoogendam, S. Klein, K. Hameeteman, M. M. B. Breteler, and W. J. Niessen. Cerebellum segmentation in MRI using atlas registration and local multi-scale image descriptors. In *Proceedings of IEEE International Symposium on Biomedical Imaging: Macro to Nano*, 2009.

R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative. LEAP: learning embeddings for atlas propagation. *NeuroImage*, 49(2):1316–1325, 2010.

Chapter 4

Rate of hippocampal atrophy and cognitive decline

This chapter is based on:
Tom den Heijer[1], Fedde van der Lijn[1], Peter J. Koudstaal, Albert Hofman, Aad van der Lugt, Gabriel P. Krestin, Wiro J. Niessen and Monique M.B. Breteler. "A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline". *Brain*, vol. 133, no. 4, pp. 1163-1172, 2010.

Abstract

Hippocampal atrophy is frequently observed on magnetic resonance images from patients with Alzheimer's disease and persons with mild cognitive impairment. Even in asymptomatic elderly, a small hippocampal volume on magnetic resonance imaging is a risk factor for developing Alzheimer's disease. However, not everyone with a small hippocampus develops dementia. With the increased interest in the use of sequential magnetic resonance images as potential surrogate biomarkers of the disease process, it has also been shown that the rate of hippocampal atrophy is higher in persons with Alzheimer's disease compared to those with mild cognitive impairment and the healthy elderly. Whether a higher rate of hippocampal atrophy also predicts Alzheimer's disease or subtle cognitive decline in non-demented elderly is unknown. We examine these associations in a group of 518 elderly (age 60-90 years, 50% female), taken from the population-based Rotterdam Scan Study. A magnetic resonance imaging examination was performed at baseline in 1995-96, and was repeated in 1999-2000 (in 244 persons) and in 2006 (in 185 persons). Using automated segmentation procedures, we assessed hippocampal volumes on all magnetic resonance imaging scans. All persons were free of dementia at baseline and followed over time for cognitive decline and incident dementia. Persons had four repeated neuropsychological tests at the research centre over a 10-year period. We also continuously monitored the medical records of all 518 participants for incident dementia. During a total follow-up of 4360 person-years, (mean 8.4, range 0.1-11.3), 50 people developed incident dementia (36 had Alzheimer's disease). We found an increased risk to develop incident dementia per standard deviation faster rate of decline in hippocampal volume [left hippocampus 1.6 (95% confidence interval 1.2-2.3, right hippocampus 1.6 (95% confidence interval 1.2-2.1)]. Furthermore, decline in hippocampal volume predicted onset of clinical dementia when corrected for baseline hippocampal volume. In people who remained free of dementia during the whole follow-up period, we found that decline in hippocampal volume paralleled, and preceded, specific decline in delayed word recall. No associations were found in this sample between rate of hippocampal atrophy, Mini Mental State Examination and tests of executive function. Our results suggest that rate of hippocampal atrophy is an early marker of incipient memory decline and dementia, and could be of additional value as a surrogate biomarker of dementia compared with a single hippocampal volume measurement.

[1] Both authors contributed equally.

4.1 Introduction

One of the major challenges in Alzheimer research is to identify persons in the earliest phase of the disease, as these persons may enter clinical trials [Sonnen et al., 2008]. Persons with so-called mild cognitive impairment (MCI) [Petersen et al., 1999] have subjective complaints and have a high conversion risk to develop clinical Alzheimer's disease (AD) [Palmer et al., 2008b]. However, half of the Alzheimer patients have never reported subjective memory complaints before diagnosis and would not come to attention of medical care or specialized memory clinics [Palmer et al., 2008a]. Therefore, objective biomarkers of the onset of the disease process, irrespective of complaints or cognitive symptoms are necessary. Moreover, association studies of early biomarkers and genetic and environmental factors could give insight in the pathogenesis and etiology of the disease.

With the increased use of magnetic resonance imaging (MRI)-based biomarkers, attention has focused on the medial temporal lobe as this region is clearly and early affected by the specific Alzheimer neuropathology [Braak and Braak, 1997]. A large number of studies found smaller hippocampal volumes on MRI in patients with AD or MCI compared with healthy controls [Convit et al., 1995, Fox et al., 1996, Jack et al., 1992, 1997]. We have previously shown that even in elderly without cognitive symptoms or complaints, a small hippocampal volume on MRI predicts AD [den Heijer et al., 2006]. However, we also showed that a large portion of people with smaller hippocampal volumes on MRI do not develop dementia.

To further improve prediction of AD or cognitive decline, follow-up brain imaging may distinguish persons with a small, yet stable volume, from those with a declining volume due to a neurodegenerative process. Longitudinal MRI scanning of the hippocampus has been performed before in a few studies. In a set of young patients with familial AD, hippocampal volume change was found to be an earlier and better predictor compared with a single volume measurement [Ridha et al., 2006]. In the elderly, rates of hippocampal atrophy on MRI were found to be higher in cases with Alzheimer and MCI compared with controls [Du et al., 2004, Jack et al., 2004]. However, in another follow-up study of three years among 27 elderly AD patients, longitudinal MRI hippocampal data did not improve diagnostic accuracy over a single volume measurement [Laakso et al., 2000].

In the current study we examine whether decline in hippocampal volume on MRI is associated with cognitive decline and incident clinical dementia, as measured in the Rotterdam Scan Study during a ten-year follow-up.

4.2 Methods

4.2.1 Setting and participants

The Rotterdam Scan Study is a large population-based cohort study among nondemented elderly in the Netherlands with baseline examinations from 1995 to 1996 [Breteler, 2000]. For details on selection criteria and differences between partici-

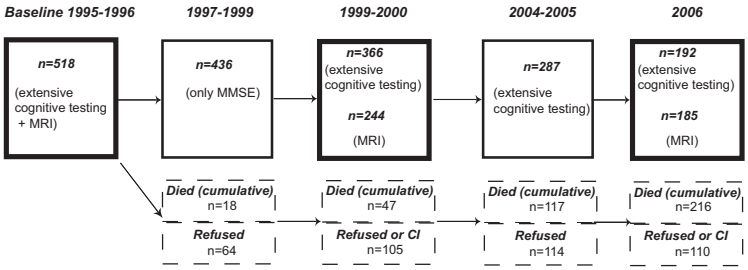


Figure 4.1: *Time frame of the study cohort (N = 518 at baseline in 1995-1996). Outlined in dark boxes are examination rounds which included MRI scanning. Number of participants at those rounds are given for the MRI part and neuropsychological examinations. At each examination round, in dashed boxes are given the number of participants that had died before we could invite them for participation (cumulative over time), and the number that refused to visit the research center, or had a contraindication (CI) for MRI scanning.*

pants and non-participants we refer to de Leeuw et al. [1999]. In 1995-1996, 518 non-demented elderly (age 60-90, 50% female) underwent, among other examinations, three-dimensional (3D) brain MRI scanning and cognitive testing. After these baseline examinations, there were four different examination rounds with cognitive testing and two follow-up brain MRI scans within a time frame of 10 years (most recent examination in 2006). Figure 4.1 shows the time frame of the study cohort and number of participants at each examination round. All participants had given written informed consent after complete description of the study. The medical ethics committee of Erasmus MC approved the study protocol.

4.2.2 MR imaging

At examinations in 1995 to 1996, in 1999-2000, and in 2006, the whole brain was imaged using a 1.5 Tesla MRI unit. The sequence of the first two MRI examinations was a custom made 3D Half-Fourier Acquisition Single-Shot Turbo Spin Echo (HASTE) sequence (inversion time 440 ms, repetition time 2800 ms, 128 contiguous sagittal slices of 1.25 mm, acquisition matrix 192×256, field of view 256×256 mm). Two HASTE modules were sequentially acquired after the inversion pulse (effective echo time of 29 ms and 440 ms) of which the first was used for volumetric assessments of the hippocampus [den Heijer et al., 2003].

Due to the availability of newer MRI techniques and a new MR scanner, the third examination (in 2006), was performed with a 3D T1 weighted sequence (3D-FSPGR-IR-T1 scan (inversion time 400 ms, repetition time 14.8 ms, time to echo 2.8 ms, 96 axial slices of 1.6 mm interpolated to 192 slices of 0.8 mm, acquisition matrix 416×256, field of view 250×250 mm)). At baseline, we performed manual segmentations of the hippocampus [den Heijer et al., 2003]. However, due to the labor-intensive nature of such measurements and the risk of errors in reproducibility, we

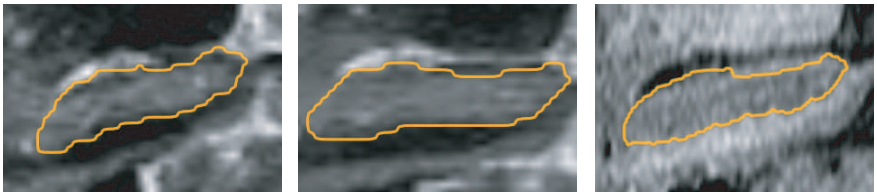


Figure 4.2: *An example of the automated segmentation of the hippocampus in a sagittal plane for all three examinations of the same subject. From left to right are shown a baseline HASTE MRI scan (absolute volume 4.0 ml), the first follow-up HASTE MRI scan (absolute volume 3.9 ml), and the second follow-up T1-weighted scan (absolute volume 3.1 ml).*

developed an automated method to segment the hippocampus on sequential MRI scans.

4.2.3 Automated segmentation of the hippocampus

The hippocampus was segmented using a variant of a previously described segmentation method [van der Lijn et al., 2008], adapted to segment the hippocampus on sequential MR images. The two most important components of this method are a statistical intensity model and a spatial probability map. The intensity model describes the typical intensities of the hippocampus and the background. The spatial probability map contains for every voxel the probability that it is part of the hippocampus.

For the segmentation of the baseline examination, the intensity model was learned from a subset of 20 scans selected from the baseline population, in which the hippocampus was manually segmented by two trained observers. The spatial probability map was created by non-rigidly registering the same 20 labeled images to the unlabeled target image, deforming the manual segmentations, and averaging them. These components were then combined to obtain the segmentation.

The first follow-up was acquired with the same MRI sequence as the baseline examination. Consequently, the same intensity model could be used for these images. For the scans at the first follow-up, the spatial probability map was obtained by first non-rigidly registering the baseline to the follow-up image, and subsequently deforming the baseline probability map.

For the second follow-up the baseline intensity model could not be used, since those images were obtained with a T1-weighted sequence. Therefore, we created a new training set by manually segmenting the hippocampus in 18 scans, which were acquired with the same scanner and sequence as the second follow-up examinations. The spatial probability map was again created by registering the baseline image to the images at second follow-up and deforming the baseline spatial probability map.

The results of all automated hippocampal segmentations were visually inspected by a single rater (T.d.H.) who was blinded for cognitive status. In a number of cases

the segmented area included the entorhinal cortex or extended to the collateral sulcus. In these instances, the segmentation was manually corrected using FSLView[2] and volumes were recalculated. Manual correction was necessary in 212 (11%) of all 1894 hippocampus assessments. Example results of the automated segmentation for all three time points are shown in Figure 4.2.

To estimate the effect of the different MRI sequences on the volume measurements, 8 elderly subjects were scanned with the 3D-HASTE sequence and 3D-T1-weighted sequence within a three-month period. The HASTE images were segmented according to the baseline procedure. The T1-weighted images were segmented with the intensity model of the second follow-up. For these latter images, the probability map was created by registering the HASTE image and deforming its probability map. The hippocampal volumes derived from both MRI sequences within this short follow-up of three months should be approximately identical. There was indeed a strong correlation between the two hippocampal volume measurements (Pearson $r = 0.97$ ($p < 0.001$)). However, the mean total hippocampus volume measured in the HASTE images was 6.13 ± 0.98 ml, versus 5.28 ± 0.98 ml in the T1 images.

Because of this systematic undersegmentation in the T1-weighted images, we could not infer absolute volume decline over the ten-year follow-up. Ranking of subjects according to rate of hippocampal decline was however possible. Therefore, we transformed all hippocampal volumes to Z-scores (individual volume-population mean/standard deviation) at each time point. By definition, the average Z-scores at each point time were zero with a standard deviation of one. If a person declined in hippocampal volume more rapidly than his or her peers he or she would decline in Z-score over time.

4.2.4 Rate of hippocampal atrophy

The decline in hippocampal volume was modeled using a linear random-effects model. This approach uses all available hippocampal data, and accounts for within-person correlation over time, which results in increased statistical power for estimating effects [Diggle et al., 1994]. We used PROC Mixed models (Statistical SAS 9.1, PROC MIXED) to model hippocampal volume decline. Taking the population with at least one repeated hippocampal volume measurement, we used hippocampal volumes at baseline and follow-up as outcome variable, and follow-up time from baseline as independent variable. The estimated fixed effect and the individual random effects were added to obtain estimated slopes of the individual Z-score declines in hippocampal volume.

4.2.5 Incident dementia

None of the 518 participants of the cohort had dementia at baseline. We followed the cohort for incident dementia with a strict protocol [den Heijer et al., 2006, Ruitenberg et al., 2001, Vermeer et al., 2003]. Briefly, participants were cognitively

[2] www.fmrib.ox.ac.uk/fsl/

screened at follow-up visits (1997-1999, 1999-2000, 2004-2005, 2006) with the Mini Mental State Examination (MMSE) and the Geriatric Mental Schedule (GMS). When screened positive, they were assessed with the Cambridge Examination for Mental Disorders of the Elderly interview. Participants who were then thought to have dementia were examined by a neurologist and underwent additional neuropsychological testing by a neuropsychologist. The number of participants that could be examined in person at the follow-up visits was 436 in 1997-1999, 366 in 1999-2000, 287 in 2004-2005, and 192 in 2006.

To avoid missing incident dementia cases among the persons who did not come to the research center, we also continuously monitored the medical records of all 518 participants at the general practitioners' offices to obtain information on diagnosed dementia and other major morbidity or death. This was done as part of the strict follow-up protocol of the Rotterdam Study [Ott et al., 1998]. Research assistants regularly checked the medical records of all participants, and medical discharge letters of memory clinics, neurologists, and the Regional Institute for Ambulatory Mental Health Care were given to the research physician. Follow-up of medical record information until January 1st 2006 was complete for more than 99% of the cohort. A diagnosis of dementia and subtype was made by a panel that consisted of a neurologist, neuropsychologist and research physician with the use of standard international criteria (McKhann et al., 1984; Román et al., 1993). The onset of dementia was defined as the date on which clinical symptoms allowed the diagnosis of dementia. Duration of follow-up for each participant was calculated from baseline examination until death, diagnosis of dementia, or the end of follow-up, whichever came first.

4.2.6 Cognitive decline

In addition to the short cognitive screening for dementia with MMSE and GMS, persons underwent extensive neuropsychological testing [Prins et al., 2005]. In short, we assessed memory function by means of a 15-word verbal learning test. The sum of words recalled at three trials was used to define immediate recall. After having done other cognitive tests in 15 minutes, a delayed recall phase was introduced. We assessed executive function with the Stroop test (part three interference), and the Letter-Digit Substitution Task (LDST). All tests were done at the research center during the follow-up examination rounds, except in 1997-1999 at which time only MMSE screening was done. For each person, we therefore had a maximum of five MMSE scores (including baseline) and four other neuropsychological test scores over the ten-year follow-up period.

To determine cognitive decline in any of these tests, we used a linear random-effects model similar to that used for hippocampal volume decline. After excluding persons with incident dementia (as they could have extreme cognitive tests results or unreliable data), we used PROC Mixed models to model cognitive decline. Cognitive decliners in each separate neuropsychological test were defined as having an individual rate of decline one standard deviation faster than the average cognitive decline.

4.2.7 Confounders

The following variables were used as potential confounders: age at baseline, sex, educational level, total gray matter volume, and white matter lesion volume. The latter two volumes were measured on the 3D MRI scans using an automated segmentation procedure volume [Ikram et al., 2008]. Hippocampal volume was calculated as per mille of the total intracranial volume. This was measured using an automated method on the baseline MRI [Ikram et al., 2008].

4.2.8 Data analysis

We used Cox' proportional hazards models to quantify the association between hippocampal volume (baseline and decline) and risk of incident dementia. Hazard ratios of dementia were calculated per standard deviation decrease of baseline Z-score or standard deviation of Z-score of hippocampal decline. For the analyses on decline in hippocampal volume, we used only incident dementia cases with a baseline and a first follow-up scan before the clinical diagnosis was made.

First, we investigated overall dementia, then separately incident Alzheimer dementia. ANCOVA was used to compare the means of MRI and cognitive variables between persons with and without dementia. We used logistic regression to quantify the association between hippocampal volume (baseline and decline) and cognitive decline. To assess whether hippocampal decline could precede cognitive decline we also separately investigated the decline in hippocampal volume from baseline to first follow-up with cognitive decline after the first follow-up MRI as dependent variable. Finally, we used ANCOVA to compare the means of all MRI volumes and individual cognitive scores between the cognitive decliners and non-decliners.

4.3 Results

Baseline characteristics and characteristics at time of follow-up MRI scanning are shown in Table 4.1. From the 518 persons at baseline, 128 had three MRI scans, 173 had two MRI scans, and 217 had a baseline MRI scan only. Average time between the first and the last MRI scan was 10.4 years (range 9.7-11.1). Persons who did not have any of the two follow-up MRI examinations after baseline (N = 217) were in general older (at baseline 4.5 years, $p < 0.001$), had a lower MMSE (-0.5 points, $p < 0.001$), had a lower baseline hippocampal Z-score (-0.22, $p = 0.02$ both for left and right hippocampal volume), but were similar with respect to sex distribution compared with persons who had at least one follow-up MRI (N = 301).

The absolute decline in left hippocampal volume from the baseline to first follow-up scan was 0.52% per year and in right hippocampal volume 0.51% per year. As expected due to the systematic undersegmentation of the hippocampus on the third MRI scan, the absolute volume decline from the first to second follow-up MRI was higher (1.7% per year for the left hippocampus and 1.6% per year for the right hippocampus).

Table 4.1: *Characteristics of the study population.*

	Baseline MRI (1995-1996)	First MRI follow-up (1999-2000)	Second MRI follow-up (2006)
Number	518	244	185
Follow-up durations, years	0.0	3.4 (0.3)	10.4 (0.4)
Age, years	73.5 (7.9)	75.3 (7.9)	79.3 (6.2)
Sex, % female	50	51	55
Only primary education, %	30	28	22
Hippocampus left, ‰ of ICV	2.69 (0.29)	2.66 (0.27)	2.39 (0.37)
Hippocampus right, ‰ of ICV	2.76 (0.28)	2.73 (0.26)	2.48 (0.35)
Total gray matter, % of ICV	46.6 (4.3)	47.7 (4.0)	45.7 (3.8)
White matter lesion volume, ml	15.1 (16.8)	11.7 (15.2)	13.2 (13.8)
Total intracranial volume, ml	1129.5 (116.2)	1126.4 (114.7)	1121.5 (115.7)
Incident dementia*, *N*	-	24	50
MMSE, score	27.7 (2.1)	27.6 (2.1)	26.5 (3.4)

Numbers are mean (SD) unless otherwise specified.
*Total number of incident dementia cases at the date on which the last MRI scan was made in the specified period.

During a follow-up period of 4360 person-years, (mean 8.4, range 0.1-11.3), 50 persons developed incident dementia (36 of them had Alzheimer's disease). Of these 50 persons, 21 were diagnosed based on the in-person screening and 29 were diagnosed based on medical information. Similar to findings previously reported in this dataset, baseline hippocampal volumes were associated with risk of dementia (age, sex, and education adjusted HR for SD decrease in left hippocampus 2.3 (95% CI 1.7-3.1) and for the right hippocampus 2.0 (95% CI 1.5-2.6)). There was no difference in effect size of the association between the hippocampus and Alzheimer's disease or non-Alzheimer dementia (data not shown).

Of the 50 persons with incident dementia, 13 had two MRI scans before their clinical diagnosis. We found that one-standard deviation faster decline in hippocampal volume was associated with a higher risk to develop dementia (age, sex, and education adjusted HR for SD decrease in left hippocampus 1.6 (95% CI 1.2-2.3) and for the right hippocampus 1.6 (95% CI 1.2-2.1)). After adjustment for baseline hippocampal volumes, the risk associated with decline in hippocampal volume remained.

Baseline total gray matter volume was not significantly associated with decline in hippocampal volume (Pearson r = -0.07, p = 0.31 for decline in left hippocampal volume and r = -0.08, p = 0.21 for decline in right hippocampal volume). More extensive white matter lesions at baseline were associated with faster decline in hip-

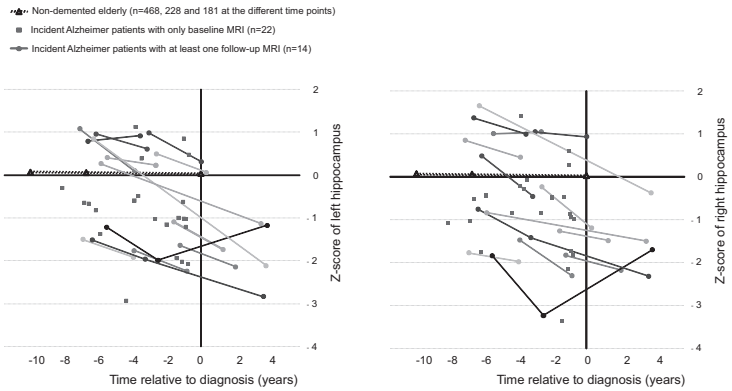


Figure 4.3: *Z-scores of hippocampal volumes relative to time to clinical diagnosis of Alzheimer's disease. A Z-score of -1.0 signifies a hippocampal volume -1.0 SD below the population mean. As a reference, the Z-scores of hippocampal of all persons who did not develop dementia over the ten year follow-up are displayed. Persons with incident Alzheimer's disease with follow-up MRI are color coded, so that the same individuals can be identified for the left and right trajectory of hippocampal decline.*

pocampal volume (Pearson correlation coefficient -0.16, p=0.02 for decline in left hippocampal volume and -0.16, p=0.01 for decline in right hippocampal volume). Adjusting the associations between decline in hippocampal volume and risk of dementia for baseline gray matter volume and white matter lesions did not change the relations (data not shown).

Figure 4.3 shows the Z-scores of the hippocampus relative to the specific clinical diagnosis of Alzheimer dementia. The majority of persons with incident Alzheimer dementia (24 of the 36 (67%) for the left hippocampus and 27 of the 36 (75%) for the right hippocampus) had years before clinical diagnosis already a Z-score of hippocampal volume below population average (i.e < 0).

Table 4.2 shows the average absolute Z-scores of hippocampal, gray matter volume and white matter lesions, at baseline and first MRI follow-up, in persons with and without incident dementia. The cognitive test scores at baseline and first follow-up of the persons with incident dementia are shown in Table 4.3.

We then investigated the association between decline in hippocampal volume and cognitive decline in the cohort of persons who remained dementia free (Table 4.4). There were 414 persons without incident dementia who had at least one repeated cognitive test after baseline. Of this group 283 subjects had one or more follow-up hippocampal volume measurements.

Those who declined faster in hippocampal volume had a statistically significant faster decline in delayed memory recall. We similarly found that a standard deviation faster decline in hippocampal volume from first to second scan predicted the

Table 4.2: *MRI characteristics of persons with incident dementia versus those without incident dementia.**

	Baseline MRI (1995-1996)	First MRI follow-up (1999-2000)	Second MRI follow-up (2006)
	Left hippocampus (Z-score)		
No incident dementia (*N*=228)	0.11 (0.91)	0.06 (0.96)	-0.01 (0.14)
Incident dementia (*N* =13)	-0.39 (1.25)	-0.77 (1.27)†	-0.12 (0.09)†
	Right hippocampus (Z-score)		
No incident dementia (*N*=228)	0.10 (0.90)	0.06 (0.93)	-0.01 (0.14)
Incident dementia (*N*=13)	-0.22 (1.37)	-0.63 (1.58)†	-0.13 (0.12)†
	Grey matter volume (Z-score)		
No incident dementia (*N*=228)	0.01 (1.01)	0.01 (1.00)	-0.00 (0.17)
Incident dementia (*N*=13)	-0.08 (0.89)	-0.15 (0.95)	-0.03 (0.13)
	White matter lesion volume (Z-score)		
No incident dementia (*N*=228)	0.02 (1.01)	0.01 (1.00)	-0.00 (0.18)
Incident dementia (*N*=13)	-0.34 (0.91)	-0.24 (0.95)	0.03 (0.22)

Adjusted means (SD) are given. Adjusted for age, sex, educational level, and total intracranial volume.
*Only persons with incident dementia were included that were diagnosed after the first follow-up MRI. Persons who developed incident dementia between baseline and first follow-up (*N*=3) were excluded as well as persons who had no first follow-up MRI (*N*=274) leaving a total of 241 persons for analyses.
†$p < 0.05$ compared with no incident dementia.

risk to develop decline in delayed memory after the second scan (odds ratio for left hippocampus 1.39 (95% CI 0.96-2.00 and for the right hippocampus 1.55 (95% CI 1.02-2.35). We found a borderline statistically significant association between decline in right hippocampal volume and risk to decline in LDST (odds ratio 1.34 (95% CI 0.98-1.82), p=0.06)). Additional adjustment for baseline total gray matter volume and white matter lesions did not change any of the associations.

We repeated the analyses for persons who had all three MRI scans available (*N* = 128) with Z-scores based on these subjects only. Similar associations were found, though with lower statistical significance, between Z-score hippocampal decline and decline in delayed recall (odds ratio per SD decline in Z-score 1.54 (95% CI 0.98-2.42, p=0.06) for the left hippocampus and 1.35 (95% CI 0.85-2.14, p=0.19) for the right hippocampus).

When we focused on the cognitive profiles of persons who had a decline in delayed recall, we found that they performed worse at any time in several other cognitive tests including MMSE (Table 4.5). They had specifically lower hippocampal

Table 4.3: *Cognitive characteristics of persons with incident dementia versus those without incident dementia.**

	Baseline MRI (1995-1996)	First MRI follow-up (1999-2000)	Score decline per year
		MMSE	
No incident dementia (*N*=241)	28.0 (1.8)	27.8 (1.8)	-0.06 (0.57)
Incident dementia (*N*=13)	26.7 (2.6)†	26.0 (2.8)†	-0.23 (0.71)
		Immediate word recall	
No incident dementia (*N*=241)	20.8 (5.3)	24.9 (4.8)	1.2 (1.4)
Incident dementia (*N*=13)	18.7 (4.3)	18.0 (5.8)†	-0.2 (1.6)†
		Delayed word recall	
No incident dementia (*N*=241)	6.5 (2.7)	7.8 (3.0)	0.4 (0.8)
Incident dementia (*N*=13)	4.6 (2.3)†	4.4 (2.3)†	-0.1 (0.6)†
		Stroop	
No incident dementia (*N*=241)	58.2 (24.9)	53.3 (20.6)	-1.2 (6.1)
Incident dementia (*N*=13)	57.4 (18.7)	56.0 (23.3)	-0.4 (3.1)
		LDST**	
No incident dementia (*N*=241)	27.8 (6.6)	27.1 (7.0)	-0.2 (1.3)
Incident dementia (*N*=13)	26.0 (7.4)	20.5 (9.3)	-1.6 (1.2)

Adjusted means (SD) are given. Adjusted for age, sex, and educational level.
*Only persons with incident dementia were included that were diagnosed after the first follow-up MRI. Persons with incident dementia between baseline and first follow-up (*N*=3) were excluded as well as persons who had no first follow-up MRI.
**Letter-Digit Substitution Task.
†$p < 0.05$ compared with no incident dementia.

volumes, but no total gray matter volume reduction (Table 4.6). Also at the second follow-up MRI, they had increased white matter lesion load.

4.4 Discussion

In this large population-based cohort study, we found that decline in hippocampal volume on MRI was associated with an increased risk to develop dementia or cognitive decline, particularly decline in delayed recall.

Before discussing these findings, several limitations of our study need to be addressed. Firstly, although we had a large cohort, few people developed dementia. At baseline, there was a natural selection of relatively healthy elderly willing to undergo MRI and other examinations. However, if anything, this healthy participator

Table 4.4: *Association between decline in hippocampal volume and cognitive decline in persons without incident dementia (N=283).**

Risk of decline* in				
MMSE	**Word recall**	**Delayed recall**	**Stroop**	**LDST****
		Left hippocampus		
1.26 (0.91-1.73)	1.10 (0.77-1.56)	1.40 (1.00-1.97)	1.07 (0.72-1.59)	1.37 (0.96-1.96)
		Right hippocampus		
1.24 (0.90-1.72)	1.11 (0.77-1.61)	1.66 (1.18-2.33)	1.28 (0.87-1.89)	1.35 (0.94-1.94)

Numbers are odds ratios (95% confidence interval) per SD decrease in baseline hippocampal volume and hippocampal volume decline. Adjustments were made for age, sex, educational level and total intracranial volume.
*Cognitive decline was defined as having a rate of decline one SD faster than average.
**Letter-Digit Substitution Task.

bias will lead to inconclusive findings and not to the present significant results.

Secondly, within persons who developed incident dementia only 13 of them had a repeated MRI scan before clinical diagnosis. On the other hand, the advantage of our population approach, compared with studies in clinical settings, is that we did include patients when they were in a very early phase of the dementia. Also, the large amount of neuropsychological data in persons who remained dementia free allowed us to examine subtle cognitive decline in association with (change in) hippocampal volumes.

Thirdly, although in the Rotterdam Study we have a thorough medical follow-up of all participants including those who refused in-person examination, we cannot fully exclude the possibility that we miss some persons with mild dementia who have no contact with the medical system. However, as we currently in the analyses include these persons in the no-dementia group this would only reduce the possibility of finding an association.

Fourthly, at invitation for follow-up MRI scans those who were older were less likely to participate. These older persons had the smallest hippocampal volumes at baseline. What we were now able to visualize as hippocampal decline is probably an underestimation of the true decline that we could have measured if we had the opportunity to have MRI images of everyone, including those who died in between MRI examinations. When computing the rate of decline in hippocampus between the baseline and first follow-up MRI scan we found a decline of 0.5% per year, which is comparable to other studies [Barnes et al., 2008].

Table 4.5: *Cognitive characteristics of persons without decline in delayed recall (N=343) versus those with decline in delayed recall (N=71).**

	Baseline MRI (1995-1996)	1997-1999	First MRI follow-up (1999-2000)	2004-2005	Second MRI follow-up (2006)
			MMSE		
Non-decliners (*N*=343)	28.2 (1.7)	27.9 (1.7)	28.0 (1.6)	27.9 (1.6)	27.0 (2.5)
Decliners (*N*=71)	27.1 (2.0)†	27.1 (1.9)†	27.0 (2.4)†	26.7 (2.9)†	23.9 (4.1)†
			Immediate word recall (total number of words)		
Non-decliners (*N*=343)	21.4 (4.9)	-	25.3 (5.3)	21.7 (7.8)	23.5 (6.1)
Decliners (*N*=71)	16.1 (3.7)†	-	19.7 (3.8)†	15.0 (4.5)†	17.7 (5.1)†
			Delayed word recall (total number of words)		
Non-decliners (*N*=343)	6.8 (2.2)	-	8.2 (2.6)	7.1 (2.6)	7.6 (3.0)
Decliners (*N*=71)	3.2 (1.2)†	-	4.5 (1.8)†	3.7 (1.5)†	3.1 (1.8)†
			Stroop (total seconds)		
Non-decliners (*N*=343)	56.8 (21.3)	-	52.8 (20.8)	57.2 (21.9)	61.0 (25.3)
Decliners (*N*=71)	61.4 (29.6)	-	60.0 (28.9)†	66.0 (33.1)†	69.8 (30.8)
			***LDST** (total number of correctly substituted)*		
Non-decliners (n=343)	28.2 (6.8)	-	28.0 (6.8)	26.7 (7.2)	26.8 (7.4)
Decliners (n=71)	25.7 (6.5)†	-	23.9 (6.9)†	24.4 (6.8)	23.2 (6.5)†

Displayed are age, sex and education adjusted means in groups (SD). Numbers given are at baseline.
*Cognitive decline in delayed recall was defined as having a decline one SD faster than average. Persons with incident dementia were excluded from these analyses.
**Letter-Digit Substitution Task.
†$p < 0.05$ for comparison between decliners and non-decliners.

Table 4.6: *MRI characteristics of persons without decline in delayed recall (N=343) versus those with decline in delayed recall (N=71).**

	Baseline MRI (1995-1996)	First MRI follow-up (1999-2000)	Second MRI follow-up (2006)
	Left hippocampus (Z-score)		
Non-decliners (*N*=343)	0.04 (0.98)	0.06 (0.99)	0.06 (0.98)
Decliners (*N*=71)	-0.17 (1.09)	-0.34 (1.02)†	-0.43 (1.06)†
	Right hippocampus (Z-score)		
Non-decliners (*N*=343)	0.03 (1.00)	0.06 (1.00)	0.06 (0.97)
Decliners (*N*=71)	-0.13 (0.98)	-0.35 (0.88)†	-0.45 (1.12)†
	Grey matter volume (Z-score)		
Non-decliners (*N*=343)	-0.01 (0.99)	0.03 (0.99)	0.02 (0.98)
Decliners (*N*=71)	0.04 (1.03)	-0.21 (1.08)	-0.11 (1.16)
	White matter lesion volume (Z-score)		
Non-decliners (*N*=343)	-0.03 (0.93)	-0.03 (0.90)	-0.06 (0.92)
Decliners (*N*=71)	0.13 (1.24)	0.19 (1.39)	0.47 (1.40)†

Displayed are age, sex, education and total intracranial volume adjusted means in groups (SD). Numbers given are at baseline.
*Cognitive decline in delayed recall was defined as having a decline one SD faster than average. Persons with incident dementia were excluded from these analyses.
†$p < 0.05$ for comparison between decliners and non-decliners.

Finally, during the follow-up of the MRI techniques inevitably had changed and improved, and in 2006 we used a different 3D MRI sequence than in 1999 and 1995. There was a high correlation of volume measurements in a set of persons who underwent both MRI sequences shortly after each other. However, absolute hippocampal volumes were systematically underestimated in the last MRI sequence. However, because of the systematic nature of this bias, ranking of persons according to the severity of their decline could be used. We used Z-scores to describe a person's hippocampal volume in the distribution at each point in time.

Previous studies have shown that the hippocampus on MRI is atrophied in patients with AD [Fox et al., 1996, Horn et al., 1996], patients with MCI [Devanand et al., 2007, Du et al., 2001], and in cognitively healthy elderly destined to develop AD [den Heijer et al., 2006]. We found that the majority of patients who developed dementia had a smaller baseline hippocampal volume years before their clinical diagnosis than persons who remained free of dementia. This is in line with clinical studies showing that approximately 80-90% of established Alzheimer patients have a small hippocampal volume [Colliot et al., 2008, Ridha et al., 2007]. Pathological validation studies have shown that hippocampal atrophy on MRI correlates with

the specific AD neuropathology [Bobinski et al., 2000, Gosche et al., 2002].

A longitudinal study in patients who carried an autosomal dominant mutation for AD found that a decline in hippocampal volume could be detected 5 years before the clinical diagnosis [Ridha et al., 2006]. In the elderly, follow-up studies in patients with AD and MCI have shown approximately 2-4 times faster rate of decline in hippocampal volume than in healthy controls [Jack et al., 2000, Laakso et al., 2000, Morra et al., 2009, Wang et al., 2003]. In these studies MRI scans were made when clinical diagnosis was already made. We showed that decline in hippocampus is strongly associated with risk to develop dementia, also in a general population setting.

Although decline in hippocampal volume in addition to a single measurement was predictive of AD, it was not always observed in subjects with incident AD, suggesting that within a clinical diagnosis of AD heterogeneity exists in its pathological substrate. Furthermore, hippocampal atrophy on MRI has also been described in patients with other dementia types such as frontotemporal dementia [Barnes et al., 2006], and vascular dementia [Kril et al., 2002]. Combining hippocampal volumes with volumes of other regions, such as the thalamus [de Jong et al., 2008] or changes assessed with other imaging techniques, such as diffusion tensor imaging or PET, might further improve identification of presymptomatic persons with AD.

In addition to the analyses on clinical dementia, we also showed that in persons who remained free of dementia over ten years a decline in hippocampal volume paralleled, or even could precede subtle cognitive decline, particularly in delayed memory function. No association was found with executive function tasks, although a borderline statistically significant association was found between hippocampal decline and decline in LDST. In this task though, not only executive function is tested, as remembering letters connected with digits may improve performance of this task.

In a smaller setting [Stoub et al., 2010], and a sample mixed with AD, MCI and healthy controls [Mungas et al., 2005], rate of hippocampal decline was similarly found to be associated with memory function. These findings support the notion that subtle delayed memory decline with hippocampal volume decline can be observed long before a clinical diagnosis of dementia is made. We showed that persons who had a decline in delayed memory had similarly low test scores on executive function and a faster decline in MMSE score, suggesting that they are en route to develop dementia.

However, not all persons with hippocampal decline or cognitive decline will eventually develop dementia. This could be either due to the fact they die before developing significant cognitive decline, or maybe because hippocampal volume decline can be part of healthy aging. A pathological study suggested differences in neuronal loss within the specific subregions of the hippocampus between normal aging and AD [West et al., 1994]. With recently introduced high resolution 7T MRI scanners, in-vivo subregion imaging of the hippocampus has become feasible, which could potentially allow discrimination of normal aging from Alzheimer's disease [Thomas et al., 2008].

References

J. Barnes, J. L. Whitwell, C. Frost, K. A. Josephs, M. Rossor, and N. C. Fox. Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration. *Arch Neurol*, 63(10):1434–1439, 2006.

J. Barnes, J. Foster, R. G. Boyes, T. Pepple, E. K. Moore, J. M. Schott, C. Frost, R. I. Scahill, and N. C. Fox. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage*, 40(4):1655–1671, 2008.

M. Bobinski, M. J. de Leon, J. Wegiel, S. Desanti, A. Convit, L. A. S. Louis, H. Rusinek, and H. M. Wisniewski. The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. *Neuroscience*, 95(3):721–725, 2000.

H. Braak and E. Braak. Frequency of stages of Alzheimer-related lesions in different age categories. *Neurobiol Aging*, 18(4):351–357, 1997.

M. M. Breteler. Vascular involvement in cognitive decline and dementia. epidemiologic evidence from the Rotterdam Study and the Rotterdam Scan Study. *Ann N Y Acad Sci*, 903:457–465, 2000.

O. Colliot, G. Chételat, M. Chupin, B. Desgranges, B. Magnin, H. Benali, B. Dubois, L. Garnero, F. Eustache, and S. Lehéricy. Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. *Radiology*, 248(1):194–201, 2008.

A. Convit, M. J. de Leon, C. Tarshish, S. D. Santi, A. Kluger, H. Rusinek, and A. E. George. Hippocampal volume losses in minimally impaired elderly. *Lancet*, 345 (8944):266, 1995.

L. W. de Jong, K. van der Hiele, I. M. Veer, J. J. Houwing, R. G. J. Westendorp, E. L. E. M. Bollen, P. W. de Bruin, H. A. M. Middelkoop, M. A. van Buchem, and J. van der Grond. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain*, 131(Pt 12):3277–3285, 2008.

F. E. de Leeuw, J. C. de Groot, M. Oudkerk, J. C. Witteman, A. Hofman, J. van Gijn, and M. M. Breteler. A follow-up study of blood pressure and cerebral white matter lesions. *Ann Neurol*, 46(6):827–833, 1999.

T. den Heijer, S. E. Vermeer, R. Clarke, M. Oudkerk, P. J. Koudstaal, A. Hofman, and M. M. B. Breteler. Homocysteine and brain atrophy on MRI of non-demented elderly. *Brain*, 126(Pt 1):170–175, 2003.

T. den Heijer, M. I. Geerlings, F. E. Hoebeek, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler. Use of hippocampal and amygdalar volumes on magnetic

resonance imaging to predict dementia in cognitively intact elderly people. *Arch Gen Psychiatry*, 63(1):57–62, 2006.

D. P. Devanand, G. Pradhaban, X. Liu, A. Khandji, S. D. Santi, S. Segal, H. Rusinek, G. H. Pelton, L. S. Honig, R. Mayeux, Y. Stern, M. H. Tabert, and M. J. de Leon. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology*, 68(11):828–836, 2007.

P. J. Diggle, K.-Y. Liang, and S. L. Zeger. *Analysis of Longitudinal Data*. Oxford: Clarendon Press, 1994.

A. T. Du, N. Schuff, D. Amend, M. P. Laakso, Y. Y. Hsu, W. J. Jagust, K. Yaffe, J. H. Kramer, B. Reed, D. Norman, H. C. Chui, and M. W. Weiner. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *J Neurol Neurosurg Psychiatry*, 71(4):441–447, 2001.

A. T. Du, N. Schuff, J. H. Kramer, S. Ganzer, X. P. Zhu, W. J. Jagust, B. L. Miller, B. R. Reed, D. Mungas, K. Yaffe, H. C. Chui, and M. W. Weiner. Higher atrophy rate of entorhinal cortex than hippocampus in AD. *Neurology*, 62(3):422–427, 2004.

N. C. Fox, E. K. Warrington, P. A. Freeborough, P. Hartikainen, A. M. Kennedy, J. M. Stevens, and M. N. Rossor. Presymptomatic hippocampal atrophy in Alzheimer's disease. a longitudinal MRI study. *Brain*, 119 (Pt 6):2001–2007, 1996.

K. M. Gosche, J. A. Mortimer, C. D. Smith, W. R. Markesbery, and D. A. Snowdon. Hippocampal volume as an index of Alzheimer neuropathology: findings from the nun study. *Neurology*, 58(10):1476–1482, 2002.

R. Horn, B. Ostertun, M. Fric, L. Solymosi, A. Steudel, and H. J. Möller. Atrophy of hippocampus in patients with Alzheimer's disease and other diseases with memory impairment. *Dementia*, 7(4):182–186, 1996.

M. A. Ikram, H. A. Vrooman, M. W. Vernooij, F. van der Lijn, A. Hofman, A. van der Lugt, W. J. Niessen, and M. M. B. Breteler. Brain tissue volumes in the general elderly population. the Rotterdam Scan Study. *Neurobiol. Aging*, 29(6):882–890, 2008.

C. R. Jack, R. C. Petersen, P. C. O'Brien, and E. G. Tangalos. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology*, 42(1):183–188, 1992.

C. R. Jack, R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O'Brien, E. G. Tangalos, G. E. Smith, R. J. Ivnik, and E. Kokmen. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, 49(3):786–794, 1997.

C. R. Jack, R. C. Petersen, Y. Xu, P. C. O'Brien, G. E. Smith, R. J. Ivnik, B. F. Boeve, E. G. Tangalos, and E. Kokmen. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. *Neurology*, 55(4):484–489, 2000.

C. R. Jack, M. M. Shiung, J. L. Gunter, P. C. O'Brien, S. D. Weigand, D. S. Knopman, B. F. Boeve, R. J. Ivnik, G. E. Smith, R. H. Cha, E. G. Tangalos, and R. C. Petersen. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*, 62(4):591–600, 2004.

J. J. Kril, S. Patel, A. J. Harding, and G. M. Halliday. Patients with vascular dementia due to microvascular pathology have significant hippocampal neuronal loss. *J Neurol Neurosurg Psychiatry*, 72(6):747–751, 2002.

M. P. Laakso, M. Lehtovirta, K. Partanen, P. J. Riekkinen, and H. Soininen. Hippocampus in Alzheimer's disease: a 3-year follow-up MRI study. *Biol Psychiatry*, 47(6):557–561, 2000.

J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, A. W. Toga, C. R. Jack, N. Schuff, M. W. Weiner, P. M. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. *NeuroImage*, 45(1 Suppl):S3–15, 2009.

D. Mungas, D. Harvey, B. R. Reed, W. J. Jagust, C. DeCarli, L. Beckett, W. J. Mack, J. H. Kramer, M. W. Weiner, N. Schuff, and H. C. Chui. Longitudinal volumetric MRI change and rate of cognitive decline. *Neurology*, 65(4):565–571, 2005.

A. Ott, M. M. Breteler, F. van Harskamp, T. Stijnen, and A. Hofman. Incidence and risk of dementia. the Rotterdam Study. *Am J Epidemiol*, 147(6):574–580, 1998.

K. Palmer, L. Bäckman, B. Winblad, and L. Fratiglioni. Early symptoms and signs of cognitive deficits might not always be detectable in persons who develop Alzheimer's disease. *Int Psychogeriatr*, 20(2):252–258, 2008a.

K. Palmer, L. Bäckman, B. Winblad, and L. Fratiglioni. Mild cognitive impairment in the general population: occurrence and progression to Alzheimer disease. *Am J Geriatr Psychiatry*, 16(7):603–611, 2008b.

R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*, 56(3):303–308, 1999.

N. D. Prins, E. J. van Dijk, T. den Heijer, S. E. Vermeer, J. Jolles, P. J. Koudstaal, A. Hofman, and M. M. B. Breteler. Cerebral small-vessel disease and decline in information processing speed, executive function and memory. *Brain*, 128(Pt 9): 2034–2041, 2005.

B. H. Ridha, J. Barnes, J. W. Bartlett, A. Godbolt, T. Pepple, M. N. Rossor, and N. C. Fox. Tracking atrophy progression in familial Alzheimer's disease: a serial MRI study. *Lancet Neurol*, 5(10):828–834, 2006.

B. H. Ridha, J. Barnes, L. A. van de Pol, J. M. Schott, R. G. Boyes, M. M. Siddique, M. N. Rossor, P. Scheltens, and N. C. Fox. Application of automated medial temporal lobe atrophy scale to Alzheimer disease. *Arch Neurol*, 64(6):849–854, 2007.

A. Ruitenberg, A. Ott, J. C. van Swieten, A. Hofman, and M. M. Breteler. Incidence of dementia: does gender make a difference? *Neurobiol Aging*, 22(4):575–580, 2001.

J. A. Sonnen, K. S. Montine, J. F. Quinn, J. A. Kaye, J. C. S. Breitner, and T. J. Montine. Biomarkers for cognitive impairment and dementia in elderly people. *Lancet Neurol*, 7(8):704–714, 2008.

T. R. Stoub, E. J. Rogalski, S. Leurgans, D. A. Bennett, and L. deToledo Morrell. Rate of entorhinal and hippocampal atrophy in incipient and mild ad: Relation to memory function. *Neurobiol Aging*, 31(7):1089–1098, 2010.

B. P. Thomas, E. B. Welch, B. D. Niederhauser, W. O. Whetsell, A. W. Anderson, J. C. Gore, M. J. Avison, and J. L. Creasy. High-resolution 7T MRI of the human hippocampus in vivo. *J Magn Reson Imaging*, 28(5):1266–1272, 2008.

F. van der Lijn, T. den Heijer, M. M. B. Breteler, and W. J. Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage*, 43(4):708–720, 2008.

S. E. Vermeer, N. D. Prins, T. den Heijer, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler. Silent brain infarcts and the risk of dementia and cognitive decline. *N Engl J Med*, 348(13):1215–1222, 2003.

L. Wang, J. S. Swank, I. E. Glick, M. H. Gado, M. I. Miller, J. C. Morris, and J. G. Csernansky. Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. *NeuroImage*, 20(2):667–682, 2003.

M. J. West, P. D. Coleman, D. G. Flood, and J. C. Troncoso. Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer's disease. *Lancet*, 344(8925):769–772, 1994.

Chapter 5

Baseline predictors of the rate of hippocampal atrophy

This chapter is based on:
Tom den Heijer[1], Fedde van der Lijn[1], M. Arfan Ikram, Peter J. Koudstaal, Aad van der Lugt, Gabriel P. Krestin, Henri A. Vrooman, Albert Hofman, Wiro J. Niessen and Monique M.B. Breteler. "Baseline predictors of hippocampus decline on MRI over a ten-year follow-up". *Submitted*.

Abstract

Decline of hippocampal volume on magnetic resonance images (MRI) may be considered as a surrogate biomarker of accumulating Alzheimer pathology. We studied potential risk factors for decline of hippocampal volume on MRI in the prospective population-based Rotterdam Scan Study. At baseline, 518 elderly were included, and the cohort was re-examined in 1999 and in 2006. In total 301 persons had at least two 3D-MRI scans to assess decline in hippocampal volume. Persons carrying the APOE ε4 allele had lower hippocampal volumes than persons with the ε3ε3 genotype, but the rate of decline was not influenced by APOE genotype. In persons who did not use antihypertensive treatment, both a high (>90 mmHg) or a low (<70 mmHg) diastolic blood pressure were associated with a faster decline in hippocampal volume. Also, white matter lesions on baseline MRI were associated with a faster decline. Vascular factors are associated with rate of hippocampal atrophy, suggesting that accumulating Alzheimer pathology in the hippocampus could be partly due to vascular risk factors.

[1] Both authors contributed equally.

5.1 Introduction

The majority of patients with Alzheimer's disease (AD) or mild cognitive impairment (MCI) have significant hippocampal atrophy on brain MRI compared with healthy elderly [Convit et al., 1997, Jack et al., 2004, Petersen et al., 2000]. The severity of hippocampal atrophy on MRI in vivo correlates postmortem with the Braak stage [Jagust et al., 2008], the extent of hippocampal neuronal loss [Bobinski et al., 2000], tangle burden [Csernansky et al., 2004] and β-amyloid plaques [Burton et al., 2009].

Given its clinical and pathological correlate, decline in hippocampal volume on MRI may serve as surrogate biomarker of accumulating Alzheimer pathology. Recent clinical trials with potential disease-modifying effects have used the rate of hippocampal atrophy on MRI as a separate outcome measure, in addition to clinical outcome [Fox et al., 2005, Jack et al., 2008, Schmidt et al., 2008].

Few studies have investigated risk factors for decline in hippocampal volume on MRI. The presence of the risk allele APOE ε4 was found to be associated with increased rate of hippocampal atrophy on MRI in Alzheimer patients [Morra et al., 2009, Schuff et al., 2009], in patients with MCI [Morra et al., 2009, van de Pol et al., 2007], but inconsistently in non-demented elderly [Morra et al., 2009, Schuff et al., 2009].

Of particular interest are potential associations between baseline vascular factors and hippocampal decline on MRI. Vascular factors have frequently been associated with the risk to develop dementia [Breteler, 2000, de la Torre, 2002, Korczyn, 2002], and this increased risk has mostly been attributed to vascular factors leading to white matter damage [Bohnen et al., 2009, Pantoni et al., 2009], and lacunar infarcts [Saczynski et al., 2009, Vermeer et al., 2003]. However, there is also evidence that vascular risk factors could more directly lead to Alzheimer neuropathology, such as plaques and tangles [Petrovitch et al., 2000].

Previously, we found that persons with untreated high diastolic blood pressure had smaller hippocampal volumes on MRI [den Heijer et al., 2005]. Also, we [den Heijer et al., 2005] and others [de Leeuw et al., 2004] have reported a co-occurrence of severe white matter lesions and smaller hippocampal volumes on MRI. However, these studies were cross-sectional in design, leaving uncertainty on cause and effects.

In the current study, we investigate the association between baseline predictors of rate of hippocampal decline over a ten-year period in the population-based Rotterdam Scan Study. Specifically, we investigate the APOE ε4 allele, blood pressure, carotid atherosclerosis and white matter lesions on MRI.

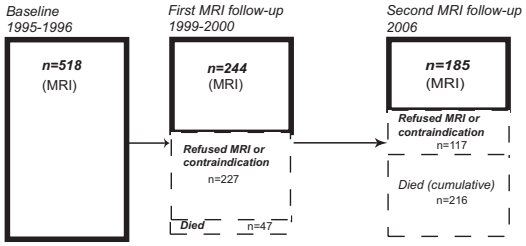


Figure 5.1: *Time frame of the study and number of participants that refused the MRI examinations, had a contraindication for MRI, or died before being invited. The size of the boxes is relative to the number of persons.*

5.2 Methods

5.2.1 Setting and participants

The Rotterdam Scan Study is a large population-based cohort study among non-demented elderly in the Netherlands with baseline examinations from 1995 to 1996 [Breteler, 2000]. This cohort was based on a subset from the original Rotterdam Study [Hofman et al., 2009]. For details on selection criteria and differences between participants and non-participants we refer to de Leeuw et al. [1999]. In 1995-1996, 518 non-demented elderly (age 60-90, 50% female) underwent three-dimensional (3D) brain MRI scanning in addition to other examinations. After these baseline examinations, there were two follow-up brain MRI examination rounds within 10 years. Figure 5.1 shows the time frame of the study cohort and the number of participants at each examination round.

As can be expected, persons who refused both follow-up MRI examinations at the research center ($N = 217$) were in general older (at baseline 4.5 years, $p < 0.001$), had a lower MMSE score (-0.8 points, $p = 0.01$), and lower baseline hippocampal volume (-0.22 in Z-score distribution, $p = 0.02$ for the left hippocampus, and $p = 0.01$ for the right hippocampus), compared with those who had at least one MRI follow-up ($N = 301$). The proportion of women, and of APOE $\varepsilon 4$ carriers did not differ between persons with, and without follow-up examinations. None of the 518 participants of the cohort had dementia at baseline. We followed the cohort for incident dementia with a strict protocol [den Heijer et al., 2006] All participants had given written informed consent after complete description of the study. The medical ethics committee of Erasmus MC approved the study protocol.

5.2.2 MR imaging and automated hippocampus segmentation

At examinations in 1995 to 1996, in 1999-2000, and in 2006, all participants had a brain MRI acquired with a 1.5 Tesla unit. The first two examinations were performed with a custom-made 3D Half-Fourier Acquisition Single-Shot Turbo Spin Echo (HASTE) sequence (inversion time 440 ms, repetition time 2800 ms, 128 con-

tiguous sagittal slices of 1.25 mm, acquisition matrix 192×256, field of view 256×256 mm). Two HASTE modules were sequentially acquired after the inversion pulse (effective echo time of 29 ms and 440 ms), of which the first was used for volumetric assessments of the hippocampus [den Heijer et al., 2003]. The third examination in 2006, was performed with a newer 1.5 Tesla scanner, and a 3D T1 weighted sequence (3D-FSPGR-IR-T1 scan (inversion time 400 ms, repetition time 14.8 ms, time to echo 2.8 ms, 96 axial slices of 1.6 mm interpolated to 192 slices of 0.8 mm, acquisition matrix 416×256, field of view 250×250 mm)).

At baseline, we had performed manual segmentations of the hippocampus [den Heijer et al., 2003]. However, because of the effort required to manually segment the follow-up scans, and the risk of reproducibility errors, we instead used an automated method to segment the hippocampus on all three sequential MR images [den Heijer et al., 2010]. More details on the segmentation can be found in Section 4.2.3 of the previous chapter.

Because of an undersegmentation in the image acquired in the third examination, we could not reliably measure the absolute rate of hippocampal decline (see Section 4.2.3). However, due to the systematic nature of the bias, ranking of subjects according to rate of hippocampal decline was possible. Therefore, we transformed all hippocampal volumes to Z-scores (individual volume-population mean divided by the standard deviation) at each time point. By definition, the average Z-scores at each point time were zero with a standard deviation of one. If a person declined in hippocampal volume more rapidly than its counterparts, he or she would decline in Z-score over time.

5.2.3 Rate of hippocampal atrophy

The decline in hippocampal volume was modeled using a linear random-effects model. This approach increases statistical power by using all available data and accounting for within-person correlation over time [Diggle et al., 1994]. The analysis was performed with PROC Mixed models (Statistical SAS 9.1, PROC MIXED). From all participants with at least one repeated hippocampal volume measurement, we used the hippocampus volumes at baseline and follow-up as outcome variable, and follow-up time as independent variable. The estimated fixed effect and the individual random effects were added to estimate slopes of the individual decline in Z-score of hippocampal volume.

5.2.4 Potential risk factors of rate of hippocampal atrophy

APOE genotyping was performed on coded DNA samples without knowledge of MRI measurements. Genotyping was available in 443 (86%) subjects out of the 518 subjects at baseline. The remainder was missing due to lack of serum samples. Blood pressure levels were measured with a random-zero sphygmomanometer. Measurements were performed twice in a seated position and the average of the two assessments was used.

Participants showed their prescribed medication to the research physician and these were coded according to the Anatomic-Therapeutic-Chemical index. Antihypertensive medications were medications in classification codes C02, C03, and C07. Hypertension was defined present when a person had either a systolic blood pressure $\geq$160 mmHg, a diastolic blood pressure $\geq$95 mmHg, or used antihypertensive medication.

Baseline white matter lesions on MRI and gray matter volume were assessed using an automated segmentation tool [Ikram et al., 2008]. Absolute volumes were divided by intracranial volume to account for head size.

At baseline, participants underwent ultrasonography of the carotid arteries [Bots et al., 1997]. The presence of atherosclerotic plaques was determined at six locations: common carotid artery, carotid bifurcation, and internal carotid artery at the left and right side and summed (range 0-6). We measured the intima-media thickness by longitudinal two-dimensional ultrasound of the anterior and posterior wall of both common carotid arteries [Bots et al., 1997]. We calculated the mean of these four locations.

5.2.5 Confounders

As potential covariates in the associations, we used body mass index (BMI), pack-years of cigarette smoked (=number of cigarettes per day X years of smoking/20), diabetes mellitus (defined present if participants reported use of oral antidiabetic-medication or insulin, or if they had a random serum glucose concentration higher than, or equal to 11.1 mmol/l). Serum total cholesterol was determined with an automated enzymatic procedure.

5.2.6 Data analysis

The relation between age and sex, and the rate of atrophy of hippocampal volume was studied using multiple linear regression. The association between APOE genotype and cross-sectional hippocampal volume was analyzed with analysis of covariance (ANCOVA) models at each MRI examination round. Persons with the APOE $\varepsilon 2\varepsilon 4$ genotype were excluded. We compared the mean Z-scores of hippocampal volume adjusted for age and sex, in persons with the APOE $\varepsilon 4$ allele, and in persons with the APOE $\varepsilon 3\varepsilon 3$ genotype. We also calculated difference in Z-scores of hippocampal volume between persons with an APOE $\varepsilon 2$ allele compared with those with the APOE $\varepsilon 3\varepsilon 3$ genotype.

We made four categories of baseline blood pressure identical to den Heijer et al. [2005] and assessed the potential associations with the rate of hippocampal atrophy in persons with, and without antihypertensive medications using ANCOVA. Furthermore, we analyzed the association between blood pressure categories and survival, to check for survival bias in the association between blood pressure and hippocampal decline.

Finally, associations between baseline white matter lesions, gray matter, carotid atherosclerosis and rate of hippocampal atrophy were analyzed with linear regres-

Table 5.1: *Characteristics of the study population.*

	Baseline MRI (1995-1996)	First MRI follow-up (1999-2000)	Second MRI follow-up (2006)
Number	518	244	185
Average follow-up time, years	0	3.4 (0.3)	10.4 (0.4)
Age, years	73.5 (7.9)	75.3 (7.9)	79.3 (6.2)
Sex, % female	50	51	55
MMSE, score	27.7 (2.1)	27.6 (2.1)	26.5 (3.4)
APOE ε4 allele, %*	32	30	32
Systolic blood pressure, mmHg	145.9 (20.6)	143.7 (22.7)	-
Diastolic blood pressure, mmHg	76.7 (11.5)	73.1 (11.1)	-
White matter lesions, % of ICV	1.4 (1.6)	-	-
Gray matter volume, % of ICV	46.6 (4.3)	-	-
Carotid plaques (range 0-6)	1.6 (1.6)	-	-
Intima-media thickness, mm	0.87 (0.14)	-	-
Hippocampus left, ‰of ICV	2.7 (0.3)	2.7 (0.3)	2.4 (0.4)
Hippocampus right, ‰of ICV	2.8 (0.3)	2.7 (0.3)	2.5 (0.4)

Values at time of MRI examination are means (standard deviation) unless specified otherwise.
*Presence of ε4 allele excluding persons with genotype ε2ε4. Available for 380 persons at baseline, 182 at first follow-up, 135 at second follow-up.

sion models. Adjustments were made for age and sex. Additionally, we adjusted for BMI, diabetes mellitus, pack-years of smoking, total cholesterol, and blood pressure (in the associations with white matter lesions).

5.3 Results

Table 5.1 shows the characteristics of the study sample at time of baseline MRI and follow-up examinations. From the 518 persons at baseline, 128 had three MRI scans, 173 had two MRI scans, and 217 had only a baseline MRI scan. Three-hundred one persons had at least one repeated MRI after baseline. Average time between the first and the third (last) MRI scan was 10.4 years (range 9.7-11.1). The decline in left hippocampal volume from the baseline to first follow-up scan was 0.52% per year, and 0.51% per year in right hippocampal volume.

Older age was associated with smaller baseline hippocampal volumes on MRI. Age at baseline was associated with a faster decline in hippocampal volume from baseline to first follow-up MRI (per 10 years of baseline age -0.03 in Z-score per year for the left hippocampus (95% CI -0.06;-0.01, p = 0.002), and -0.05 in Z score

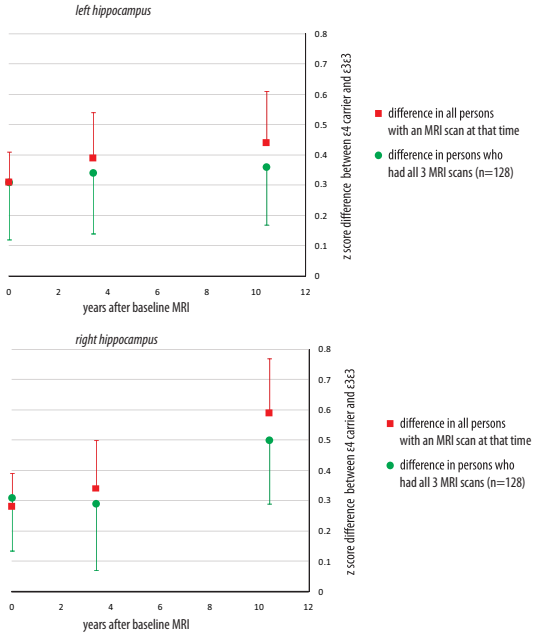


Figure 5.2: *Difference in Z-scores of left (upper panel) and right (lower panel) hippocampus with standard errors at the three MRI points between carriers of the ε4 allele compared with persons with the ε3ε3 genotype. We excluded persons with the ε2ε4 genotype. The difference in Z-score is displayed for each MRI time point using either all persons who had an MRI scan at that time in red (N=518 at baseline, N=244 at first follow-up and N=185 at second follow-up,) and in green for persons who had underwent all three sequential MRI scans (N=128).*

per year for the right hippocampus (95% CI -0.08;-0.03, $p < 0.001$)). Age at baseline did not predict decline of the hippocampus over the whole follow-up period. Men had a significantly faster decline in hippocampal volume on the right, but not on the left, compared with women (-0.01 in Z score per year (95% CI -0.02;-0.00, $p = 0.01$).

As shown in Figure 5.2, persons carrying the APOE ε4 allele had significantly lower hippocampal volumes at each MRI time point compared with persons with the ε3ε3 genotype. The results were not different when we analyzed only persons who had all three MRI scans, but statistical significance was lower due to smaller sample sizes.

Rate of hippocampal atrophy over time was slightly higher in APOE ε4 carriers but this was not statistically significant (for left hippocampus $p = 0.56$ and for right hippocampus $p = 0.11$). Persons carrying an APOE ε2 allele had no different hippocampal volume at any time compared with persons with the ε3ε3 genotype. There were only 12 incident dementia cases with repeated MRI hippocampal as-

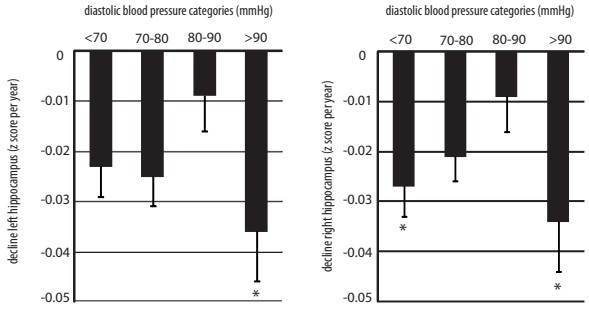


Figure 5.3: *The association between diastolic blood pressure levels and hippocampal decline on MRI in persons untreated for hypertension. Values are mean values (+standard error) of decline in hippocampal volume per year (decline in Z-score of the hippocampus per year). Values are age and sex adjusted means. *$p < 0.05$ compared with diastolic blood pressure category 80-90 mm Hg.*

sessments; therefore we could not specifically investigate potential effects of APOE $\varepsilon 4$ allele on hippocampal decline in persons with incident dementia.

Hypertension at baseline was not associated with rate of hippocampal atrophy (data not shown). However, in persons not treated with antihypertensive medications, baseline diastolic blood pressure levels were associated with hippocampal decline. Figure 5.3 shows that both a low and a high diastolic blood pressure were associated with a faster decline (for the left hippocampus the p-value of quadratic term of diastolic blood pressure was 0.03, and for the right hippocampus the p-value of the quadratic term was 0.007).

In order to rule out that survival bias influenced these results, we looked at the association between diastolic blood pressure category and survival. Within the total clinical follow-up period until 1st January, 2007, 235 persons died. We found that persons not using antihypertensive treatment in the lowest diastolic blood pressure category had the lowest risk to die in the observed follow-up period (<70 mmHg 30% died; 70-80 mmHg 35% died; 80-90 mmHg 35% died; > 90 mmHg 52% died). Additional adjustment for BMI, pack-years of smoking, diabetes mellitus and cholesterol levels did not change these associations.

White matter lesion load on MRI at baseline, but not carotid atherosclerosis, was significantly associated with a faster decline in hippocampal volume on MRI over time (Table 5.2). When we additionally adjusted these associations for blood pressure, BMI, pack-years of smoking, diabetes mellitus and cholesterol levels, these results did not change. Of the other baseline MRI measures, we found no association between total intracranial volume or total gray matter volume and hippocampal decline (data not shown).

Table 5.2: *Associations between markers of vasculopathy at baseline and rate of hippocampal volume decline on MRI (N=301)*

	Left hippocampus decline		**Right hippocampus decline**	
	Estimate (95% CI)	*p*-value	Estimate (95% CI)	*p*-value
Carotid atherosclerosis				
Intima-media thickness	0.00 (-0.01;0.01)	0.89	0.00 (-0.01;0.01)	0.80
Carotid plaques	-0.00 (-0.01;0.00)	0.56	0.00 (-0.01;0.01)	0.75
White matter lesions on MRI	-0.01 (-0.02;-0.01)	<0.001	-0.01 (-0.02;-0.00)	0.002

Values are age and sex adjusted regression coefficients per standard deviation increase in carotid atherosclerosis or white matter lesions on MRI. A negative regression coefficient indicates a faster hippocampal decline.

5.4 Discussion

In the current large population-based study with a long follow-up, we show that baseline vascular factors are associated with rate of hippocampal atrophy. Carriers of the APOE ε4 allele had at any point in time smaller hippocampal volume than non-carriers, but they did not have a faster decline in hippocampal volume.

Some limitations of our study need to be discussed. Firstly, our study subjects were relatively healthy. Especially persons who participated in follow-up MRI examinations were more likely to be healthy survivors. We found that persons with untreated high diastolic blood pressure had a higher mortality rate in the observed follow-up period. Therefore, the observed association between high blood pressure and decline in hippocampal volume is probably an underestimation of the true effect, which we only could have determined if we had the possibility to include all persons.

A second limitation is that as the MRI techniques over the follow-up improved, with a different 3D MRI sequence and scanner used in 2006 than in 1999 and 1995. There was a high correlation of volume measurements in a set of persons who underwent these 3D MRI sequences on both scanners within a short interval. However, absolute hippocampal volumes were systematically underestimated when measured from the MRI sequence used in 2006. Although ranking of persons according to severity of decline in hippocampal volume can therefore be used, we cannot infer absolute decline in volume over this ten-year period. When computing the rate of atrophy in hippocampus between the baseline and first follow-up MRI scan three years later (which were based on similar MRI sequences), we found

a decline of 0.5% per year which is comparable to other studies [Barnes et al., 2008].

Hippocampal atrophy on MRI in vivo could serve as biomarker of Alzheimer pathology [Bobinski et al., 2000, Csernansky et al., 2004, Gosche et al., 2002, Jagust et al., 2008]. Rate of hippocampal atrophy on MRI is increasingly being used as a separate outcome measure in clinical trials [Fox et al., 2005, Jack et al., 2008, Schmidt et al., 2008]. In our own population we recently found that rate of hippocampal atrophy on MRI predicted risk of dementia and decline in delayed recall [den Heijer et al., 2010]. This clinical and pathophysiological relevance underscores the importance to find risk factors of hippocampal decline on MRI.

We found that persons carrying the APOE ε4 allele had lower hippocampal volumes compared with persons with the ε3ε3 genotype, at any point of time over ten years of follow-up. Previous studies have shown effects of the APOE ε4 allele on hippocampal volume in midlife and in non-demented persons [den Heijer et al., 2002, Lind et al., 2006, Schmidt et al., 1996, Tohgi et al., 1997], suggesting that a small hippocampal volume could be an inherited trait. Alternatively, higher occurrence of Alzheimer pathology in midlife APOE ε4 carriers [Kok et al., 2009] could have led to smaller hippocampal volumes.

We found rates of atrophy in hippocampus not to be higher in APOE ε4 carriers. This is in concordance with a study in the Alzheimer Disease Neuroimaging Initiative (ADNI) showing a faster rate of atrophy of the hippocampus only in Alzheimer patients carrying the ε4 allele, yet not in healthy controls or persons with mild cognitive impairment with the ε4 allele [Schuff et al., 2009]. The number of incident Alzheimer patients in our study was too small to separately investigate the effect of the APOE ε4 allele in them.

Taken together, these data suggest that the effect of the APOE ε4 allele on hippocampal volume is persistent through life, but that only in combination with significant Alzheimer neuropathology a faster rate of hippocampal atrophy may be observed on MRI.

Vascular factors increase the risk to develop clinical Alzheimer's disease [Breteler, 2000, de la Torre, 2002, Launer et al., 2000]. Pathological series have shown that persons who died with a clinical diagnosis of Alzheimer's disease frequently have lacunar infarcts and small vessel disease [MRC CFAS, 2001]. This suggests that vascular factors through lacunar infarcts and small vessel disease cause cognitive decline in a brain with existing Alzheimer pathology [Snowdon et al., 1997]. We examined in the current study whether vascular factors could also be more directly involved by increasing Alzheimer pathology, using hippocampal decline on MRI as surrogate in-vivo biomarker.

We observed that persons with a higher baseline white matter lesions volume load had an increased rate of hippocampal atrophy. In a smaller set of patients with Alzheimer's disease, such an association has been described before [de Leeuw et al., 2006]. The association between white matter lesions and hippocampal atrophy could be due to two underlying pathophysiological mechanisms.

First, hippocampal atrophy and white matter lesions may be caused by similar risk factors such as long-term hypertension and atherosclerosis. High blood pres-

sure is a well-known risk factor for white matter lesions [van Swieten et al., 1991]. In a cross-sectional study we previously found that high diastolic blood pressure levels are associated with low hippocampal volumes [den Heijer et al., 2005]. The results presented in this report also show similar associations between high diastolic blood pressure and the rate of hippocampal atrophy.

This is in line with a study showing in cognitively intact hypertensive persons a reduced regional cerebral blood flow to the hippocampus [Dai et al., 2008]. In rats, chronic brain hypoperfusion was found to cause selective capillary abnormalities in the CA1 region of the hippocampus [Jong et al., 1999]. Midlife hypertension was also reportedly associated with hippocampal atrophy on MRI in later life [Korf et al., 2004], and extent of neurofibrillary tangles and amyloid plaques in the hippocampus [Petrovitch et al., 2000].

We additionally show in this chapter that low diastolic blood pressure levels are associated with increased rates of hippocampal atrophy. These persons may have chronic cerebral hypoperfusion, especially when cerebral autoregulation cannot compensate for low systemic blood pressure [Duschek and Schandry, 2007]. In our population this association was not driven by people with the lowest diastolic blood pressures being the frailest. On the contrary, persons with the lowest diastolic blood pressure levels had the lowest mortality over the follow-up period.

Although white matter lesions and hippocampal atrophy could be caused by a concomitant risk factor such as hypertension, adjusting the association between white matter lesions and rate of hippocampal atrophy for blood pressure and other potential risk factors did not change the association. Also, white matter lesions have not a fully overlap in risk factor profile with decline in hippocampal volume, as carotid atherosclerosis was not associated with rate of hippocampal atrophy but is strongly associated with white matter lesions [de Leeuw et al., 2000].

A second explanation for the association between white matter lesions and hippocampal atrophy may be that one causes the occurrence of the other. Damage to the white matter could disrupt normal connections of the efferent projections of hippocampus. This in turn could cause retrograde (Wallerian) loss of hippocampal axons. However, hippocampal neuronal damage due to accumulating Alzheimer pathology could also lead to white matter disintegrity in connecting fiber tracts such as the cingulum [Villain et al., 2008] and fornix [Mielke et al., 2009].

References

J. Barnes, J. Foster, R. G. Boyes, T. Pepple, E. K. Moore, J. M. Schott, C. Frost, R. I. Scahill, and N. C. Fox. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage*, 40(4):1655–1671, 2008.

M. Bobinski, M. J. de Leon, J. Wegiel, S. Desanti, A. Convit, L. A. S. Louis, H. Rusinek, and H. M. Wisniewski. The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. *Neuroscience*, 95(3):721–725, 2000.

N. I. Bohnen, M. L. T. M. Müller, H. Kuwabara, G. M. Constantine, and S. A. Studenski. Age-associated leukoaraiosis and cortical cholinergic deafferentation. *Neurology*, 72(16):1411–1416, 2009.

M. L. Bots, A. W. Hoes, P. J. Koudstaal, A. Hofman, and D. E. Grobbee. Common carotid intima-media thickness and risk of stroke and myocardial infarction: the Rotterdam Study. *Circulation*, 96(5):1432–1437, 1997.

M. M. Breteler. Vascular involvement in cognitive decline and dementia. epidemiologic evidence from the Rotterdam Study and the Rotterdam Scan Study. *Ann N Y Acad Sci*, 903:457–465, 2000.

E. J. Burton, R. Barber, E. B. Mukaetova-Ladinska, J. Robson, R. H. Perry, E. Jaros, R. N. Kalaria, and J. T. O'Brien. Medial temporal lobe atrophy on MRI differentiates Alzheimer's disease from dementia with lewy bodies and vascular cognitive impairment: a prospective study with pathological verification of diagnosis. *Brain*, 132(Pt 1):195–203, 2009.

A. Convit, M. J. D. Leon, C. Tarshish, S. D. Santi, W. Tsui, H. Rusinek, and A. George. Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiol Aging*, 18(2):131–138, 1997.

J. G. Csernansky, J. Hamstra, L. Wang, D. McKeel, J. L. Price, M. Gado, and J. C. Morris. Correlations between antemortem hippocampal volume and postmortem neuropathology in ad subjects. *Alzheimer Dis Assoc Disord*, 18(4):190–195, 2004.

W. Dai, O. L. Lopez, O. T. Carmichael, J. T. Becker, L. H. Kuller, and H. M. Gach. Abnormal regional cerebral blood flow in cognitively normal elderly subjects with hypertension. *Stroke*, 39(2):349–354, 2008.

J. C. de la Torre. Alzheimer disease as a vascular disorder: nosological evidence. *Stroke*, 33(4):1152–1162, 2002.

F. E. de Leeuw, J. C. de Groot, M. Oudkerk, J. C. Witteman, A. Hofman, J. van Gijn, and M. M. Breteler. A follow-up study of blood pressure and cerebral white matter lesions. *Ann Neurol*, 46(6):827–833, 1999.

F. E. de Leeuw, J. C. de Groot, M. L. Bots, J. C. Witteman, M. Oudkerk, A. Hofman, J. van Gijn, and M. M. Breteler. Carotid atherosclerosis and cerebral white matter lesions in a population based magnetic resonance imaging study. *J Neurol*, 247 (4):291–296, 2000.

F.-E. de Leeuw, F. Barkhof, and P. Scheltens. White matter lesions and hippocampal atrophy in Alzheimer's disease. *Neurology*, 62(2):310–312, 2004.

F.-E. de Leeuw, E. Korf, F. Barkhof, and P. Scheltens. White matter lesions are associated with progression of medial temporal lobe atrophy in Alzheimer disease. *Stroke*, 37(9):2248–2252, 2006.

T. den Heijer, M. Oudkerk, L. J. Launer, C. M. van Duijn, A. Hofman, and M. M. B. Breteler. Hippocampal, amygdalar, and global brain atrophy in different apolipoprotein e genotypes. *Neurology*, 59(5):746–748, 2002.

T. den Heijer, S. E. Vermeer, R. Clarke, M. Oudkerk, P. J. Koudstaal, A. Hofman, and M. M. B. Breteler. Homocysteine and brain atrophy on MRI of non-demented elderly. *Brain*, 126(Pt 1):170–175, 2003.

T. den Heijer, L. J. Launer, N. D. Prins, E. J. van Dijk, S. E. Vermeer, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler. Association between blood pressure, white matter lesions, and atrophy of the medial temporal lobe. *Neurology*, 64(2):263–267, 2005.

T. den Heijer, M. I. Geerlings, F. E. Hoebeek, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler. Use of hippocampal and amygdalar volumes on magnetic resonance imaging to predict dementia in cognitively intact elderly people. *Arch Gen Psychiatry*, 63(1):57–62, 2006.

T. den Heijer, F. van der Lijn, P. J. Koudstaal, A. Hofman, A. van der Lugt, G. P. Krestin, W. J. Niessen, and M. M. B. Breteler. A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain*, 133(Pt 4):1163–1172, 2010.

P. J. Diggle, K.-Y. Liang, and S. L. Zeger. *Analysis of Longitudinal Data*. Oxford: Clarendon Press, 1994.

S. Duschek and R. Schandry. Reduced brain perfusion and cognitive performance due to constitutional hypotension. *Clin Auton Res*, 17(2):69–76, 2007.

N. C. Fox, R. S. Black, S. Gilman, M. N. Rossor, S. G. Griffith, L. Jenkins, M. Koller, and A. N1792(Q. S-21)-201 Study. Effects of abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology*, 64(9):1563–1572, 2005.

K. M. Gosche, J. A. Mortimer, C. D. Smith, W. R. Markesbery, and D. A. Snowdon. Hippocampal volume as an index of Alzheimer neuropathology: findings from the Nun Study. *Neurology*, 58(10):1476–1482, 2002.

A. Hofman, M. M. B. Breteler, C. M. van Duijn, H. L. A. Janssen, G. P. Krestin, E. J. Kuipers, B. H. C. Stricker, H. Tiemeier, A. G. Uitterlinden, J. R. Vingerling, and J. C. M. Witteman. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol*, 24(9):553–572, 2009.

M. A. Ikram, H. A. Vrooman, M. W. Vernooij, F. van der Lijn, A. Hofman, A. van der Lugt, W. J. Niessen, and M. M. B. Breteler. Brain tissue volumes in the general elderly population. the Rotterdam Scan Study. *Neurobiol. Aging*, 29(6):882–890, 2008.

C. R. Jack, M. M. Shiung, J. L. Gunter, P. C. O'Brien, S. D. Weigand, D. S. Knopman, B. F. Boeve, R. J. Ivnik, G. E. Smith, R. H. Cha, E. G. Tangalos, and R. C. Petersen. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*, 62(4):591–600, 2004.

C. R. Jack, R. C. Petersen, M. Grundman, S. Jin, A. Gamst, C. P. Ward, D. Sencakova, R. S. Doody, L. J. Thal, and Members of the Alzheimer's Disease Cooperative Study (ADCS). Longitudinal MRI findings from the vitamin e and donepezil treatment study for mci. *Neurobiol Aging*, 29(9):1285–1295, 2008.

W. J. Jagust, L. Zheng, D. J. Harvey, W. J. Mack, H. V. Vinters, M. W. Weiner, W. G. Ellis, C. Zarow, D. Mungas, B. R. Reed, J. H. Kramer, N. Schuff, C. DeCarli, and H. C. Chui. Neuropathological basis of magnetic resonance images in aging and dementia. *Ann Neurol*, 63(1):72–80, 2008.

G. I. D. Jong, E. Farkas, C. M. Stienstra, J. R. Plass, J. N. Keijser, J. C. de la Torre, and P. G. Luiten. Cerebral hypoperfusion yields capillary damage in the hippocampal ca1 area that correlates with spatial memory impairment. *Neuroscience*, 91(1):203–210, 1999.

E. Kok, S. Haikonen, T. Luoto, H. Huhtala, S. Goebeler, H. Haapasalo, and P. J. Karhunen. Apolipoprotein E-dependent accumulation of Alzheimer disease-related lesions begins in middle age. *Ann Neurol*, 65(6):650–657, 2009.

A. D. Korczyn. Mixed dementia–the most common cause of dementia. *Ann N Y Acad Sci*, 977:129–134, 2002.

E. S. C. Korf, L. R. White, P. Scheltens, and L. J. Launer. Midlife blood pressure and the risk of hippocampal atrophy: the honolulu asia aging study. *Hypertension*, 44 (1):29–34, 2004.

L. J. Launer, G. W. Ross, H. Petrovitch, K. Masaki, D. Foley, L. R. White, and R. J. Havlik. Midlife blood pressure and dementia: the Honolulu-Asia aging study. *Neurobiol Aging*, 21(1):49–55, 2000.

J. Lind, A. Larsson, J. Persson, M. Ingvar, L.-G. Nilsson, L. Bäckman, R. Adolfsson, M. Cruts, K. Sleegers, C. V. Broeckhoven, and L. Nyberg. Reduced hippocampal volume in non-demented carriers of the apolipoprotein e epsilon4: relation to chronological age and recognition memory. *Neurosci Lett*, 396(1):23–27, 2006.

M. M. Mielke, N. A. Kozauer, K. C. G. Chan, M. George, J. Toroney, M. Zerrate, K. Bandeen-Roche, M.-C. Wang, P. Vanzijl, J. J. Pekar, S. Mori, C. G. Lyketsos, and M. Albert. Regionally-specific diffusion tensor imaging in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, 46(1):47–55, 2009.

J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, A. W. Toga, C. R. Jack, N. Schuff, M. W. Weiner, P. M. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. *NeuroImage*, 45(1 Suppl):S3–15, 2009.

MRC CFAS. Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). *Lancet*, 357(9251):169–175, 2001.

L. Pantoni, A. Poggesi, and D. Inzitari. Cognitive decline and dementia related to cerebrovascular diseases: some evidence and concepts. *Cerebrovasc Dis*, 27 Suppl 1:191–196, 2009.

R. C. Petersen, C. R. Jack, Y. C. Xu, S. C. Waring, P. C. O'Brien, G. E. Smith, R. J. Ivnik, E. G. Tangalos, B. F. Boeve, and E. Kokmen. Memory and MRI-based hippocampal volumes in aging and AD. *Neurology*, 54(3):581–587, 2000.

H. Petrovitch, L. R. White, G. Izmirilian, G. W. Ross, R. J. Havlik, W. Markesbery, J. Nelson, D. G. Davis, J. Hardman, D. J. Foley, and L. J. Launer. Midlife blood pressure and neuritic plaques, neurofibrillary tangles, and brain weight at death: the haas. honolulu-asia aging study. *Neurobiol Aging*, 21(1):57–62, 2000.

J. S. Saczynski, S. Sigurdsson, M. K. Jonsdottir, G. Eiriksdottir, P. V. Jonsson, M. E. Garcia, O. Kjartansson, O. Lopez, M. A. van Buchem, V. Gudnason, and L. J. Launer. Cerebral infarcts and cognitive performance: importance of location and number of infarcts. *Stroke*, 40(3):677–682, 2009.

H. Schmidt, R. Schmidt, F. Fazekas, J. Semmler, P. Kapeller, B. Reinhart, and G. M. Kostner. Apolipoprotein E e4 allele in the normal elderly: neuropsychologic and brain MRI correlates. *Clin Genet*, 50(5):293–299, 1996.

R. Schmidt, S. Ropele, B. Pendl, P. Ofner, C. Enzinger, H. Schmidt, A. Berghold, M. Windisch, H. Kolassa, and F. Fazekas. Longitudinal multimodal imaging in mild to moderate Alzheimer disease: a pilot study with memantine. *J Neurol Neurosurg Psychiatry*, 79(12):1312–1317, 2008.

N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. M. Shaw, J. Q. Trojanowski, P. M. Thompson, C. R. Jack, M. W. Weiner, and Alzheimer's Disease Neuroimaging Initiative. MRI of hippocampal volume loss in early Alzheimer's disease in relation to apoe genotype and biomarkers. *Brain*, 132(Pt 4):1067–1077, 2009.

D. A. Snowdon, L. H. Greiner, J. A. Mortimer, K. P. Riley, P. A. Greiner, and W. R. Markesbery. Brain infarction and the clinical expression of Alzheimer disease. the nun study. *JAMA*, 277(10):813–817, 1997.

H. Tohgi, S. Takahashi, E. Kato, A. Homma, R. Niina, K. Sasaki, H. Yonezawa, and M. Sasaki. Reduced size of right hippocampus in 39- to 80-year-old normal subjects carrying the apolipoprotein e epsilon4 allele. *Neurosci Lett*, 236(1):21–24, 1997.

L. A. van de Pol, W. M. van der Flier, E. S. C. Korf, N. C. Fox, F. Barkhof, and P. Scheltens. Baseline predictors of rates of hippocampal atrophy in mild cognitive impairment. *Neurology*, 69(15):1491–1497, 2007.

J. C. van Swieten, G. G. Geyskes, M. M. Derix, B. M. Peeck, L. M. Ramos, J. C. van Latum, and J. van Gijn. Hypertension in the elderly is associated with white matter lesions and cognitive decline. *Ann Neurol*, 30(6):825–830, 1991.

S. E. Vermeer, N. D. Prins, T. den Heijer, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler. Silent brain infarcts and the risk of dementia and cognitive decline. *N Engl J Med*, 348(13):1215–1222, 2003.

N. Villain, B. Desgranges, F. Viader, V. de la Sayette, F. Mézenge, B. Landeau, J.-C. Baron, F. Eustache, and G. Chételat. Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in Alzheimer's disease. *J Neurosci*, 28(24):6174–6181, 2008.

Chapter 6

Regional white matter lesion measurement

This chapter is based on:
Fedde van der Lijn, Ben F. Verhaaren, M. Arfan Ikram, Stefan Klein, Henri A. Vrooman, Meike W. Vernooij, Alexander Hammers, Daniel Rueckert, Aad van der Lugt, Monique M.B. Breteler and Wiro Niessen. "Automated measurement of local white matter lesion volume". *Submitted*.

Abstract

It has been hypothesized that white matter lesions at different locations may have different etiology and clinical consequences. Several approaches for the measurement of local white matter lesion load have been proposed in the literature, most of which rely on a distinction between lesions in a periventricular region close to the ventricles and a subcortical zone further away. In this work we present a novel automated method for local white matter lesion volume quantification in magnetic resonance images. The method segments and measures the white matter lesion volume in 43 regions defined by orientation and distance with respect to the ventricles, which allows a more detailed study of lesion location. The potential of the method was demonstrated by analyzing the effect of blood pressure on the regional white matter lesion load in 490 elderly subjects taken from a longitudinal population study. The method was also compared to two commonly used techniques to assess the periventricular and subcortical lesion load. The main finding was that high blood pressure was mainly associated with lesion load in the vascular watershed area that forms the border between the periventricular and subcortical regions. This is compatible to the associations between blood pressure and periventricular/subcortical load computed for the same data, and corresponds with findings reported in the literature. However, the proposed method allows analyses with a higher spatial resolution.

6.1 Introduction

Elderly individuals often exhibit hyperintense lesions on T2-weighted or fluid attenuation inversion recovery (FLAIR) magnetic resonance images (MRI) of the brain. There has been extensive debate about whether these white matter lesions (WMLs) have a different underlying etiology or varying clinical consequences depending on their location in the brain. For example, WMLs located in the periventricular areas (PVWMLs) were found to be related to cognitive decline, whereas WMLs found in the subcortical white matter (SCWMLs) were associated with depression [de Groot et al., 2000a,b]. A histopathological examination linked SCWMLs to ischemia, and PVWMLs to non-ischemic damage to the myelin [Fazekas et al., 1993]. Finally, a three year follow-up MRI study showed a larger rate of progression of subcortical WMLs compared to periventricular WMLs [Sachdev et al., 2007].

However, other authors have argued that this distinction between PVWMLs and SCWMLs may not be very plausible biologically or clinically relevant. Several pathological studies point to a common ischemic etiology for both types of lesions (see DeCarli et al. [2005] for an overview). Furthermore, [DeCarli et al., 2005] shows that PVWML and SCWML load are highly correlated with the total WML load and each other.

Traditionally, analyses of local WML severity are based on scoring the lesion loads in the periventricular and subcortical areas. The most commonly used measurements of PVWML and SCWML loads are visual rating scales that have separate scores for the two WML categories [de Groot et al., 2000b, Fazekas et al., 1987, Scheltens et al., 1993]. A number of more recent studies have introduced automated PVWML/SCWML analyses [DeCarli et al., 2005, van der Lijn et al., 2007, Wen and Sachdev, 2004]. These methods all follow a similar strategy: WML voxels obtained from an automated lesion segmentation are labeled as periventricular if they are located within a user-defined distance from the ventricular wall. The boundary between the periventricular and subcortical zones is usually positioned in the vascular watershed area, between 3 and 13 mm from the ventricles.

A second approach to local WML analysis are regional volume measurements. This type of method provides more detailed spatial localization information by determining the WML load in regions like the separate lobes [Wen and Sachdev, 2004, Yoshita et al., 2006], or by creating lesion probability maps that show the lesion frequency per voxel location [DeCarli et al., 2005, Enzinger et al., 2006, Jongen et al., 2009, Wen and Sachdev, 2004].

All these methods involve a trade-off between spatial resolution and power to detect associations further away from the ventricles. At voxel locations away from the ventricles the probability of encountering a lesion is much smaller because WMLs tend to spread over a larger area [Jongen et al., 2009]. This reduces the sensitivity of methods based on lesion probability maps to detect an association in the subcortical region. Aggregating over the entire subcortical region solves this problem, but it reduces the spatial resolution of the analysis.

In this work, we present an automated method for regional WML measurement

that attempts to find a balance between the PVWML/SCWML distinction and a lesion probability map. It measures WML volume, obtained with an automated segmentation technique, in 43 regions defined by their distance and orientation with respect to the ventricles. The potential of the method is demonstrated by analyzing the relation between blood pressure and regional WML volume for 490 elderly subjects taken from a population-based imaging study. The results are compared with a visual rating and automated PVWML/SCWML measurements.

6.2 Materials and methods

6.2.1 Setting and participants

The study population was derived from the Rotterdam Study, a large population-base cohort study in the Netherlands that started in 1990 and investigates the prevalence, incidence, and determinants of chronic diseases in the elderly [Hofman et al., 1991, 2007, 2009]. From 1995 to 1996 we randomly selected 965 persons between 60 and 90 years of age to participate in the Rotterdam Scan Study that investigates age-related brain abnormalities on MRI [den Heijer et al., 2003]. After excluding persons who were demented or had MRI contraindications, 832 participants were invited. Among these, 563 persons gave their written informed consent and participated in the study (response 68%). Of this group, 52 persons developed claustrophobia during MRI acquisition. Twenty one datasets were unusable for analysis, leaving a total of 490 participants with complete and usable MRI data [Ikram et al., 2008]. The study protocol was approved by the medical ethics committee of the Erasmus MC, Rotterdam, the Netherlands.

6.2.2 MR imaging

The image data were acquired with a 1.5T Siemens Vision MR unit. They consisted of proton density-, T1-, and T2-weighted images (PDw, T1w and T2w), as well as a custom-made 3D Half-Fourier Acquisition Single-Shot Turbo Spin Echo (HASTE) sequence. The PDw, T1w and T2w images had a slice thickness of 5 mm, but were coregistered and resampled to the HASTE voxel size of $1.25 \times 1 \times 1$ mm. All scans were corrected for MR bias fields with N3 [Sled et al., 1998]. Non-rigid registration of a manually labeled brain mask (excluding the cerebellum) was used to measure the intracranial volume (ICV). [Ikram et al., 2008].

6.2.3 Blood pressure

In a study centre visit prior to the MRI examinations (1990-1993), the participants' blood pressure had been measured. Two measurements were performed on the right upper arm with a random-zero sphygmomanometer, and the average of both was used. The use of blood-pressure-lowering medication was assessed during a home interview. 14 subjects were excluded from the analysis, because the blood

pressure measurement was not available (1 person), or no information about the medication was known (13 persons).

6.2.4 Regional WML volume

To investigate the importance of lesion location in more detail we developed a novel method that quantifies WML volume in 43 regions defined by their distance and orientation with respect to the ventricles. The technique consists of four steps: lesion segmentation, ventricle segmentation, region segmentation, and the regional WML volume measurement.

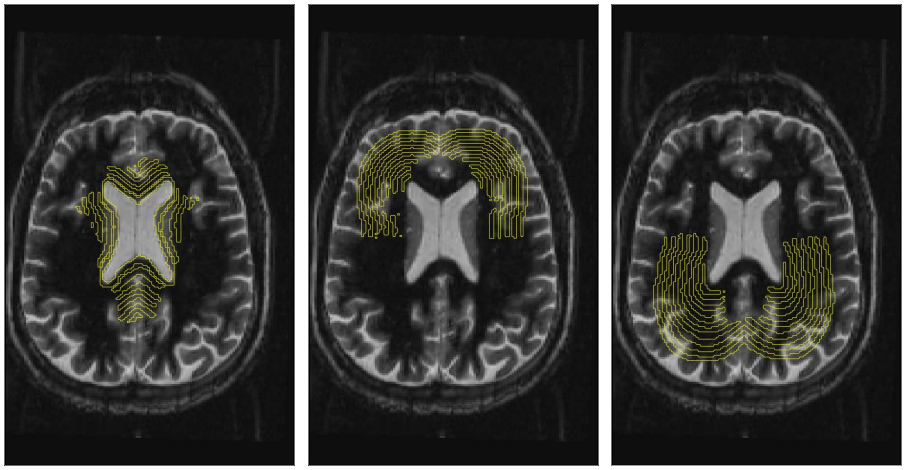
First, all images were segmented in gray matter(GM), white matter(WM), cerebrospinal fluid(CSF), and WMLs with a supervised voxel classification method [Vrooman et al., 2007]. With this technique WMLs were segmented based on their intensity on the T2w, PDw, and HASTE images. Its accuracy was tested on 9 manually segmented datasets. The average WML Dice similarity index was 0.63 and the volumetric ICC was 0.84 [Ikram et al., 2008].

As a second step we segmented the ventricles in the HASTE images, by non-rigidly registering 20 atlas images to the target images using the Elastix registration software [Klein et al., 2010]. The atlas images contained 83 manually labeled brain areas, including the ventricles [Hammers et al., 2003]. By applying the deformation obtained with the registration to the ventricle regions and averaging the results, a ventricular probability map in the coordinate system of the target image was constructed. Both this probability map, and a CSF probability image obtained with the brain tissue classifier described above were then used as input for an automated brain structure segmentation method [van der Lijn et al., 2008].

In the third step, the 43 regions were obtained by first dividing the brain in three sections that contain the three periventricular locations distinguished in the visual rating described above: the regions adjacent to the frontal horns, to the occipital horns, and to the main body of the lateral ventricles. We did not choose for a lobar subdivision, because this would spread the posterior WML volume over three lobes.

The three sections were composed from the 83 atlas regions described above. We deformed the regions to the target coordinate system and combined the results with a vote rule [Heckemann et al., 2006]. The 83 binary region labels were then merged to form the anterior, posterior, and central section. The anterior section covered the frontal lobe excluding the precentral gyrus. The posterior section contained the entire occipital lobe, a large part of the parietal lobe, and the precentral gyrus. The central section consisted of the area around the basal ganglia up to and including the insula, as well as the corpus callosum and the cingulate gyrus. The anterior part of the temporal lobe up to and including the hippocampus, the cerebellum and brain stem were excluded from the analysis as they contained very little WMLs.

These sections were then further subdivided in 15 equidistant shells around the segmented ventricles. The first boundary was located at 2 mm from the ventricles, and the last one at 28 mm. Lesions more than 28 mm from the ventricles were considered part of the last shell. Most of the WML voxels directly adjacent to the



(a) Central section (b) Anterior section (c) Posterior section

Figure 6.1: *Axial slice of an HASTE image with a subdivision in 43 regions. For illustration purposes every other boundary is not shown in the image.*

posterior and anterior horns of the ventricles fell into the corpus callosum or caudate nucleus regions, which are part of the first central shell. As a result it was not possible to accurately estimate the lesion volume in the first shell in the anterior and posterior section. We therefore considered all WML voxels up to 2 mm as part of the first central shell.

This scheme resulted in a subdivision of the brain in 43 regions (Figure 6.1). In the last step, the volumes of the segmented WMLs were measured for each region.

6.2.5 Visual periventricular and subcortical WML rating

The PVWML and SCWML load was also visually assessed for all subjects by two experienced raters out of a pool of four, using the T1w, T2w and PDw images. The PVWML load was rated separately for regions adjacent to the anterior horn, occipital horns, and the main body of the ventricles on a scale of zero to three. The three regional ratings were then summed to obtain a score between zero and nine. Subcortical lesions were counted and divided in three size categories based on their largest diameter. The total SCWML load was computed by assigning a fixed volume of 0.004, 0.1, and 0.9 ml for every lesion in each category respectively [de Groot et al., 2000b]. Weighted kappas for the PVWML grade were between 0.79 and 0.90. The SCWML inter- and intra-rater intra-class correlation coefficients (ICCs) were 0.88 and 0.95 [de Groot et al., 2000b]. These scores were computed over 100 subjects.

Table 6.1: *Characteristics of the study population.*

	N=476
Age at time of scan, years	73.4 (7.9)
Female sex, *N*(%)	244 (51.3)
Systolic blood pressure, mmHg	145.8 (20.5)
Diastolic blood pressure, mmHg	76.6 (11.5)
Average time between BP measurement and scan, years	4.7 (1.0)
Use of blood-pressure-lowering drugs, *N*(%)	145 (30.5)
Intracranial volume, ml	1127 (116)

Values are means (standard deviation) unless specified otherwise.

6.2.6 Automated periventricular and subcortical WML measurement

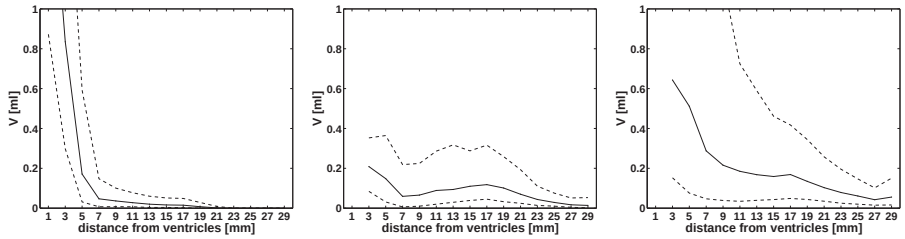
Finally, PVWML and SCWML volumes were measured with an automated method. For this measurement we divided the brain in two regions based on distance to the segmented ventricles. The periventricular and subcortical WML volumes were obtained by computing a Euclidean distance transform from the ventricle segmentation. All segmented WML voxels within 7 mm were labeled as PVWML, the rest was labeled as SCWML.

6.2.7 Data analysis

The spatial distribution of the WMLs was analyzed by computing the first quartile, median, and third quartile of the lesion volumes per section and per region. The relation between blood pressure and regional WML volume was investigated with multiple linear regression. These analyses were adjusted for age, sex, ICV, and the use of blood-pressure-lowering medication. To make the regression coefficients comparable between the regions, we converted all regional WML volumes to Z-scores by subtracting the population median and dividing by the interquartile range.

Since the regional WML volume is low for most subjects and always positive, the residuals of the linear model are not normally distributed. We therefore used a randomization test with 10000 permutations under a reduced model to compute p-values for the null-hypothesis that the slope is zero [Freedman and Lane, 1983] (cited in Anderson and Robinson [2001]). The randomization test was also used to correct for multiple comparisons. This was done by comparing the t-value associated with the slope of each regional regression to the null distribution of maximum t-values computed over all regions [Nichols and Holmes, 2002].

To compare these results to the periventricular and subcortical lesion loads, first quartiles, medians, and third quartiles were computed of the visual and automated PVWML/SCWML scores. We then computed the association between blood pres-



(a) Central section (b) Anterior section (c) Posterior section

Figure 6.2: *Median WML volumes in the consecutive shells with increasing distance to the ventricles. The dashed lines mark the first and third quartile.*

sure and the risk of severe PVWML and SCWML load with multiple logistic regression for both the visual and automated scores. Severe WMLs were defined as the upper quintile of the score or volume distribution. This approach was chosen because the discrete visual PVWML score was of a very different nature than the visual SCWML rating or the automated PVWML/SCWML volumes. All analyses were adjusted for age, sex, ICV, and the use of blood-pressure-lowering medication.

6.3 Results

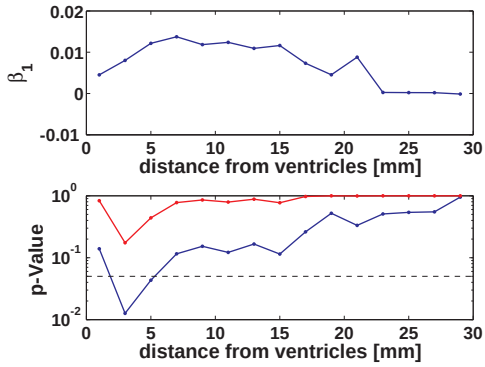
6.3.1 Subjects

The characteristics of the study population are shown in Table 6.1.

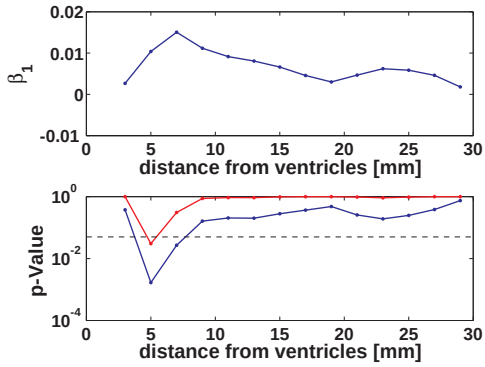
6.3.2 Regional WML analysis

The first quartile, median, and third quartile of the WML volume per section were: 0.7, 1.4, and 3.4 ml for anterior; 1.2, 3.2, and 8.9 ml for posterior; and 1.5, 3.6, and 6.9 ml for central. The excluded brain regions contained 0.00, 0.02, and 0.06 ml. Figure 6.2 shows the spatial distribution of the WMLs per section. The central section had the highest volumes in the first shell directly adjacent to the ventricles. Moving away from the ventricles the lesion load quickly dropped. The posterior and anterior section can be divided in three parts. The first part starts at 2 mm with the largest regional WML load, which then decreases until about 10 mm. Moving further away from the ventricles, the regional load stabilizes in the posterior section or rises in the anterior section. After 18 mm from the ventricles the regional load starts to decrease again until it reaches negligible quantities.

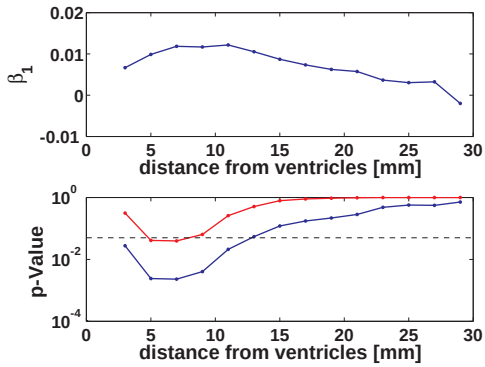
Figure 6.3 shows the strength of the relation between diastolic blood pressure and normalized regional WML volume, as a function of distance to the ventricle. The three sections show roughly the same pattern: the association strengths reach a maximum after the first few shells and then slowly drop as the ventricle distance



(a) Central section

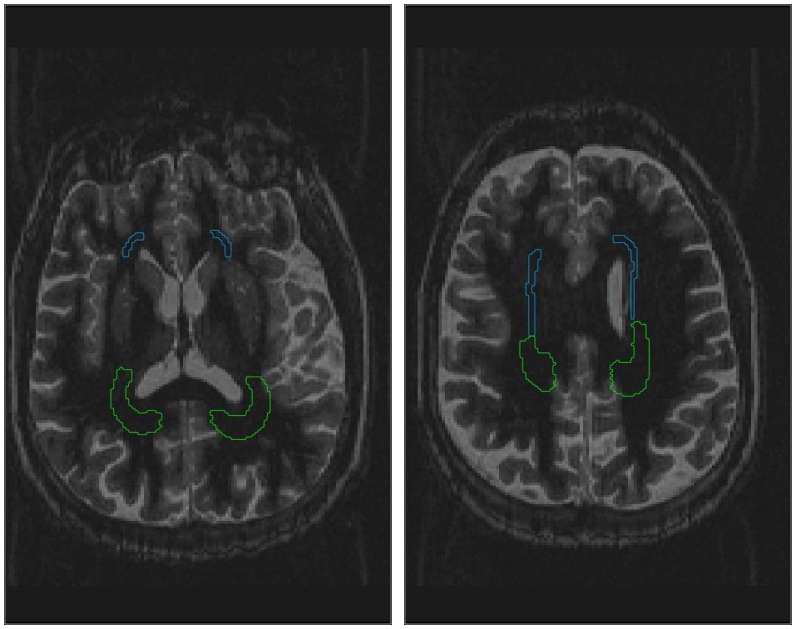


(b) Anterior section



(c) Posterior section

Figure 6.3: *Regression coefficients and p-values for the association between diastolic blood pressure and regional WML volume Z-scores, corrected for age, sex, ICV, and use of blood-pressure-lowering medication. The lower p-value curves are not corrected for multiple comparisons, the upper curves are. The dashed line represents the 0.05 level.*



(a) Posterior and anterior regions (b) Posterior and anterior regions

Figure 6.4: *Two axial slices showing the anterior and posterior shells, in which the relation between blood pressure and regional WML volume was borderline statistically significance after correction for multiple comparisons.*

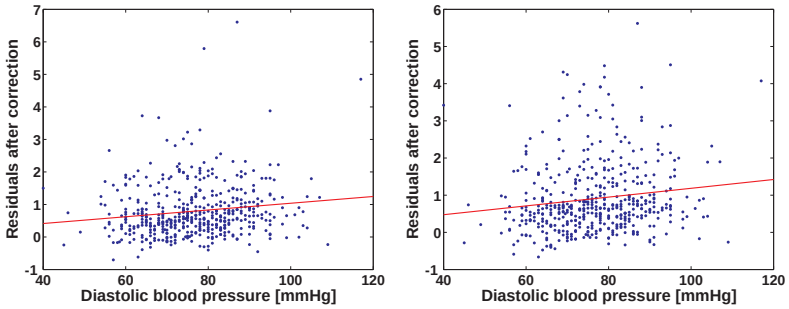
increases. p-values in the first shell are relatively high, and then drop to a minimum in the second shell. Moving away from the ventricles, the values start rising again.

Figure 6.4(a) and 6.4(b) show the anterior and posterior shells for which the relation between diastolic blood pressure and regional WML volume was borderline statistically significant after correcting for multiple comparisons. Figure 6.5(a) and 6.5(b) show the fit of the blood pressure model for two of these regions. None of the regions in the central section had a statistically significant relation between diastolic blood pressure and lesion load.

We did not find any statistically significant relation between systolic blood pressure and WML volume. In this analysis the regional association strength curves were relatively flat (data not shown).

6.3.3 Periventricular and subcortical WML analysis

The first quartile, median and second quartile of the visual PVWML rating were 1, 2, and 4. The SCWML load had a first quartile, median and third quartile of 0.0, 0.3, and 1.9 ml. The automated method measured a PVWML load of 2.0, 5.1, and 10.8 ml, and a SCWML load of 1.4, 3.8, and 8.1 ml.



(a) Anterior region 4-6 mm (b) Posterior region 6-8 mm

Figure 6.5: *Scatter plots of the diastolic blood pressure versus the regional WML volume in two regions with a statistically significant association. The y-axis shows regional Z-scores after correction for the covariates. The line represents the model computed with the regression analysis.*

Table 6.2 shows the association between blood pressure and a severe load for PVWML and SCWML as measured with the visual and automated ratings. Both methods show similar relations between higher diastolic blood pressure and severe WML load. Neither method shows an association between systolic blood pressure and the presence of a severe WML load.

6.4 Discussion

This section is divided in three parts: first the results of the regional WML analysis are interpreted and compared to the existing literature. In the second part, the method itself will be discussed and compared to alternative techniques. The section ends with the conclusion.

6.4.1 Discussion of the results

The main finding of the automated regional WML analysis is that the areas most affected by diastolic blood pressure are located between 4 and 6 mm in the anterior section, and 4 and 10 mm in the posterior section. These regions are all located in the so-called vascular watershed area around the ventricles, supporting the notion that this area is especially vulnerable to ischemia [DeCarli et al., 2005, Moody et al., 1990].

The WML load in the shell directly adjacent to the ventricles is not significantly related to diastolic blood pressure, although it contains the largest number of lesion voxels. From this result we conclude that WMLs in these areas can be found across the population regardless of blood pressure. This hypothesis is in agreement with two pathological studies [Fazekas et al., 1993, Takao et al., 1999], which suggested that small caps and pencil-thin lining around the ventricles represent normal anatomical variation.

We also did not find evidence for an association between diastolic blood pressure and WML load further away from the ventricles. The association strength and level of significance dropped gradually when moving beyond the 4 mm boundary around the ventricles. This result suggests that the presence of focal lesions outside the vascular watershed area are not heavily influenced by blood pressure.

These results were also compared to the associations found with periventricular and subcortical WMLs. With this common distinction we found a significant association between diastolic blood pressure and the risk of a severe SCWML load, and a borderline significant association with the risk of having a severe PVWML load. The associations were similar for both visual and automated PVWML/SCWML scores. These results further confirm the hypothesis that the white matter lesions in the vascular watershed area are related to blood pressure. Since the area falls in between the periventricular and subcortical zones, it contains both extensive periventricular and confluent subcortical WMLs.

Several groups have previously studied the relation between blood pressure and local WML load. For example, a meta-analysis of 10 European studies that used visual rating scales found associations between blood pressure and PVWML or SCWML load that were comparable to what was reported in this work [van Dijk et al., 2004].

Yoshita and co-authors studied the relation between hypertension and the WML load in six small regions around the ventricles. In the three smallest regions the WML volume was not significantly higher for subjects with hypertension [Yoshita et al., 2006]. Two of these were directly adjacent to the ventricles. The three regions that did show a big increase also contained parts of the vascular watershed area. These results therefore seem to confirm most of our findings.

Enzinger and co-authors investigated the difference between lesion probability maps of subjects with and without hypertension [Enzinger et al., 2006]. Their analysis did not reveal voxel locations in the watershed areas that were significantly related to hypertension. This difference with our results might be caused by a lack of power needed to do a voxel-based analysis further away from the ventricles.

6.4.2 Evaluation of the method

The proposed method has some limitations that need to be addressed. First of all, it does not measure WML load in regions defined in a template space. Most regional WML analyses [DeCarli et al., 2005, Enzinger et al., 2006, Jongen et al., 2009, Wen and Sachdev, 2004] use registration to a common template to compensate for differences in head size. Because we performed our analyses in each subject's native coordinate system, these differences might have caused additional variance on the distance measurements and weakened the associations.

On the other hand, registration to a template introduces errors. The regional segmentation used in this work is not immune to registration errors, but it allows the use of multiple atlases to compensate for this [Heckemann et al., 2006]. Multi-atlas segmentation is not possible for registration to a template space. Furthermore, WMLs can negatively influence the registration process, as severe lesions tend to

Table 6.2: *Odds ratios of baseline blood pressure for severe WML load at follow-up, as measured with the visual rating and automated measurements.*

	Visual	Automated
	PVWML	
Diastolic Blood Pressure	1.21 (0.96; 1.52)	1.15 (0.91; 1.44)
Systolic Blood Pressure	0.99 (0.87; 1.12)	0.99 (0.88; 1.12)
	SCWML	
Diastolic Blood Pressure	1.44 (1.14; 1.80)	1.32 (1.05; 1.65)
Systolic Blood Pressure	1.04 (0.93; 1.18)	1.02 (0.91; 1.15)

Values are odds ratios per 10 mmHg increase in blood pressure (95% CI). Adjusted for age, sex, intracranial volume, and the use of blood-pressure-lowering drugs.

have higher signal intensities on the HASTE scan and appear as cerebro-spinal fluid. Some papers have dealt with this problem by masking lesion areas [DeCarli et al., 2005] or basing the registration on the ventricles only [Jongen et al., 2009]. But these methods will invariably introduce some uncertainty on the deformation of the lesions. Finally, it is unclear whether the exact location of the watershed area scales with head size, or whether it is determined by its absolute distance to the closest vessels.

As can be seen in Figure 6.5, the residuals of the linear model are not normally distributed. We therefore used permutation tests to compute p-values, which do not assume normality. However, Figure 6.5 does suggest that the skewness predominantly affected the intercept and not the slope. Robust regression techniques like least absolute deviations could be used to compute a median slope, but this is beyond the scope of this work.

Finally, the spatial resolution of the proposed regional method is limited compared to WML probability maps. However, although lesion probability maps are a valuable tool to map the spatial WML distribution, we believe they are less suited to test associations with risk factors and clinical consequences. Even for relatively large regions we obtained only a few associations that were significant after correcting for multiple comparisons. This might be the reason that previous studies that used lesion maps for this purpose found very few areas of significant association [Enzinger et al., 2006, Jongen et al., 2009]. To the best of our knowledge, only one study based on probability maps has found associations not directly adjacent to the the ventricles [Taylor et al., 2003].

On the other hand, the regional analysis has a better spatial resolution than the PVWML/SCWML distinction. In this work we studied the effect of blood pressure on the WML load in the watershed area and the voxels directly adjacent to the ventricles. This would have been impossible with WML measurements in periventricular, subcortical, or lobar regions. In particular, confluent lesions in the watershed

area cannot not be accurately modeled by the simple threshold of the ventricular distance function that is used in automated PVWML/SCWML measurements. A hard distinction between two types of lesions might therefore be too crude to shed further light on the debate about the role of WML location.

6.4.3 Conclusion

This paper presented an automated method for measurement of the local white matter lesion load, based on the distance and orientation to the ventricles. The method allows a more spatially detailed analysis of the influence of lesion location on associations than most other local WML measurement techniques. We demonstrated the potential of the method by studying the relation between blood pressure and regional WML load in 490 subjects, and found that diastolic blood pressure is related to an increased lesion load in the vascular watershed area on the border between the periventricular and subcortical zones. This is compatible to the associations between blood pressure and PVWML/SCWML load computed for the same data, and is also in line with findings reported in the literature. However, the proposed method enables analyses with a higher spatial resolution.

References

M. J. Anderson and J. Robinson. Permutation tests for linear models. *Australian & New Zealand Journal of Statistics*, 43:75–88, 2001.

J. C. de Groot, F. E. de Leeuw, M. Oudkerk, A. Hofman, J. Jolles, and M. M. Breteler. Cerebral white matter lesions and depressive symptoms in elderly adults. *Arch Gen Psychiatry*, 57(11):1071–1076, 2000a.

J. C. de Groot, F. E. de Leeuw, M. Oudkerk, J. van Gijn, A. Hofman, J. Jolles, and M. M. Breteler. Cerebral white matter lesions and cognitive function: the Rotterdam Scan Study. *Ann Neurol*, 47(2):145–151, 2000b.

C. DeCarli, E. Fletcher, V. Ramey, D. Harvey, and W. J. Jagust. Anatomical mapping of white matter hyperintensities (WMH): exploring the relationships between periventricular WMH, deep WMH, and total WMH burden. *Stroke*, 36(1): 50–55, 2005.

T. den Heijer, S. E. Vermeer, E. J. van Dijk, N. D. Prins, P. J. Koudstaal, A. Hofman, and M. M. B. Breteler. Type 2 diabetes and atrophy of medial temporal lobe structures on brain MRI. *Diabetologia*, 46(12):1604–1610, 2003.

C. Enzinger, S. Smith, F. Fazekas, G. Drevin, S. Ropele, T. Nichols, T. Behrens, R. Schmidt, and P. M. Matthews. Lesion probability maps of white matter hyperintensities in elderly individuals: results of the austrian stroke prevention study. *J Neurol*, 253(8):1064–1070, 2006.

F. Fazekas, J. B. Chawluk, A. Alavi, H. I. Hurtig, and R. A. Zimmerman. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am J Roentgenol*, 149(2):351–356, 1987.

F. Fazekas, R. Kleinert, H. Offenbacher, R. Schmidt, G. Kleinert, F. Payer, H. Radner, and H. Lechner. Pathologic correlates of incidental MRI white matter signal hyperintensities. *Neurology*, 43(9):1683–1689, 1993.

D. Freedman and D. Lane. A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics*, 1(4):292–98, 1983.

A. Hammers, R. Allom, M. J. Koepp, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, and J. S. Duncan. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp*, 19(4):224–247, 2003.

R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.

A. Hofman, D. E. Grobbee, P. T. de Jong, and F. A. van den Ouweland. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol*, 7(4):403–422, 1991.

A. Hofman, M. M. B. Breteler, C. M. van Duijn, G. P. Krestin, H. A. Pols, B. H. C. Stricker, H. Tiemeier, A. G. Uitterlinden, J. R. Vingerling, and J. C. M. Witteman. The Rotterdam Study: objectives and design update. *Eur J Epidemiol*, 22(11):819–829, 2007.

A. Hofman, M. M. B. Breteler, C. M. van Duijn, H. L. A. Janssen, G. P. Krestin, E. J. Kuipers, B. H. C. Stricker, H. Tiemeier, A. G. Uitterlinden, J. R. Vingerling, and J. C. M. Witteman. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol*, 24(9):553–572, 2009.

M. A. Ikram, H. A. Vrooman, M. W. Vernooij, F. van der Lijn, A. Hofman, A. van der Lugt, W. J. Niessen, and M. M. B. Breteler. Brain tissue volumes in the general elderly population. the Rotterdam Scan Study. *Neurobiol Aging*, 29(6):882–890, 2008.

C. Jongen, J. van der Grond, P. Anbeek, M. A. Viergever, G. J. Biessels, J. P. W. Pluim, and U. D. E. S. Group. Construction of periventricular white matter hyperintensity maps by spatial normalization of the lateral ventricles. *Hum Brain Mapp*, 30 (7):2056–2062, 2009.

S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*, 29(1):196–205, 2010.

D. M. Moody, M. A. Bell, and V. R. Challa. Features of the cerebral vascular pattern that predict vulnerability to perfusion or oxygenation deficiency: an anatomic study. *AJNR Am J Neuroradiol*, 11(3):431–439, 1990.

T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*, 15(1):1–25, 2002.

P. Sachdev, W. Wen, X. Chen, and H. Brodaty. Progression of white matter hyperintensities in elderly individuals over 3 years. *Neurology*, 68(3):214–222, 2007.

P. Scheltens, F. Barkhof, D. Leys, J. P. Pruvo, J. J. Nauta, P. Vermersch, M. Steinling, and J. Valk. A semiquantative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging. *J Neurol Sci*, 114(1):7–12, 1993.

J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*, 17 (1):87–97, 1998.

M. Takao, A. Koto, N. Tanahashi, Y. Fukuuchi, M. Takagi, and S. Morinaga. Pathologic findings of silent hyperintense white matter lesions on MRI. *J Neurol Sci*, 167(2):127–131, 1999.

W. D. Taylor, J. R. MacFall, D. C. Steffens, M. E. Payne, J. M. Provenzale, and K. R. R. Krishnan. Localization of age-associated white matter hyperintensities in late-life depression. *Prog Neuropsychopharmacol Biol Psychiatry*, 27(3):539–544, 2003.

F. van der Lijn, M. W. Vernooij, M. A. Ikram, H. Vrooman, D. Rueckert, A. Hammers, M. M. B. Breteler, and W. Niessen. Automated localization of periventricular and subcortical white matter lesions. In J. P. W. J. M. Reinhardt, editor, *SPIE Medical Imaging 2007: Image Processing*, 2007.

F. van der Lijn, T. den Heijer, M. M. B. Breteler, and W. J. Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage*, 43(4):708–720, 2008.

E. J. van Dijk, M. M. B. Breteler, R. Schmidt, K. Berger, L.-G. Nilsson, M. Oudkerk, A. Pajak, S. Sans, M. de Ridder, C. Dufouil, R. Fuhrer, S. Giampaoli, L. J. Launer, A. Hofman, and C. A. S. C. A. D. E. Consortium. The association between blood pressure, hypertension, and cerebral white matter lesions: cardiovascular determinants of dementia study. *Hypertension*, 44(5):625–630, 2004.

H. A. Vrooman, C. A. Cocosco, F. van der Lijn, R. Stokking, M. A. Ikram, M. W. Vernooij, M. M. B. Breteler, and W. J. Niessen. Multi-spectral brain tissue segmentation using automatically trained k-nearest-neighbor classification. *NeuroImage*, 37(1):71–81, 2007.

W. Wen and P. Sachdev. The topography of white matter hyperintensities on brain MRI in healthy 60- to 64-year-old individuals. *NeuroImage*, 22(1):144–154, 2004.

M. Yoshita, E. Fletcher, D. Harvey, M. Ortega, O. Martinez, D. M. Mungas, B. R. Reed, and C. S. DeCarli. Extent and distribution of white matter hyperintensities in normal aging, MCI, and AD. *Neurology*, 67(12):2192–2198, 2006.

Chapter 7

Summary and discussion

The aim of this thesis was the development of accurate and robust methods for brain structure segmentation in magnetic resonance imaging (MRI) data, and the validation and use of these methods in the context of a large population imaging study. In the previous chapters, three atlas-based methods were introduced for the segmentation of brain structures in single or sequential images. The methods were applied to data from the population-based Rotterdam Scan Study to study associations of hippocampal volume and decline, and to measure regional white matter lesion volumes. This last chapter provides a summary of this work, a discussion of the main contributions, followed by a conclusion and directions for future research.

7.1 Summary

In **Chapter 2** a segmentation method was presented based on the minimization of an energy functional with intensity and prior terms that are derived from manually labeled training images. The intensity energy is implemented with a voxel classifier that learns the foreground and background intensity distributions from the training images. The prior energy consists of a spatial and regularity term. The spatial prior is obtained from a probability map created by registering the training images to the unlabeled target image, and deforming and averaging the training labels. The regularity prior energy encourages smooth segmentations. The resulting energy functional is globally minimized using graph cuts.

The method was evaluated using image data from the population-based Rotterdam Scan Study on diseases among the elderly. Two sets of images were used: a small set of 20 MR images, in which the hippocampi were manually labeled, and a larger set of 498 images, for which manual hippocampal volume measurements were available, but no segmentations. This second set was previously used in a volumetry study that found significant associations between hippocampal volume, cognitive decline, and the incidence of dementia. Cross-validation experiments with the labeled set showed similarity indices of 0.852 and 0.864 and mean surface distances of 0.40 and 0.36 mm for the left and right hippocampus. 83% of the automated segmentations of the large set were rated as 'good' by a trained observer. Also, the proposed method was used to repeat the manual hippocampal volumetry study. The automatically obtained hippocampal volumes showed significant associations with cognitive decline and dementia similar to the manually measured volumes. Finally, direct quantitative and qualitative comparisons showed that the proposed method outperformed a multi-atlas-based segmentation method.

Chapter 3 described a novel method for brain structure segmentation in MR images that combines information about a structure's location and its appearance. The spatial model is implemented by registering multiple atlas images to the unlabeled target image and creating a spatial probability map. The structure's appearance is modeled by a classifier based on Gaussian scale-space features. These components are combined with a regularization term in a Bayesian framework that is globally optimized using graph cuts. The incorporation of an appearance model enables the method to segment structures with complex intensity distributions and increases

its robustness against errors in the spatial probability map.

The method was tested in leave-one-out experiments on two datasets acquired with different MR sequences, in which the hippocampus and cerebellum were manually segmented by an expert. Furthermore, the method was compared to three other segmentation techniques based on atlas registration, atlas registration plus an intensity model, and Freesurfer, which were applied to the same data. Results showed that the atlas&appearance-based method produced accurate results with mean Dice similarity indices of 0.95 for the cerebellum, and 0.87 for the hippocampus. This was significantly better than the atlas-based method and comparable to the atlas&intensity-based method, but the atlas&appearance-based method is more widely applicable and robust. The results were compared to Freesurfer by computing correlations between the automated and manual volume measurements. These were found to be very similar for both methods.

Hippocampal atrophy is frequently observed on MR images from patients with Alzheimer's disease (AD) and persons with mild cognitive impairment (MCI). Even in asymptomatic elderly, a small hippocampal volume on an MR image is a risk factor for developing AD. However, not everyone with a small hippocampus develops dementia. It has been shown that the rate of hippocampal atrophy is higher in persons with AD, compared to those with MCI and the healthy elderly. **Chapter 4 and 5** focused on the use of hippocampal decline measured in sequential MR images as potential surrogate biomarkers of the disease process.

Whether a higher rate of hippocampal atrophy also predicts AD or subtle cognitive decline in non-demented elderly is unknown. In **Chapter 4** these associations were examined in a group of 518 elderly (age 60-90 years, 50% female), taken from the Rotterdam Scan Study. MR imaging was performed in 1995-96, and repeated in 1999-2000 (in 244 persons) and in 2006 (in 185 persons). Using an automated atlas&intensity-based segmentation method, hippocampal volumes were assessed on all MR scans. All persons were free of dementia at baseline and followed over time for cognitive decline and incident dementia. Persons had four repeated neuropsychological tests at the research center over a ten-year period. The medical records of all 518 participants were also continuously monitored for incident dementia.

During a total follow-up time of 4360 person-years, (mean 8.4, range 0.1-11.3), 50 people developed incident dementia (of whow 36 had AD). We found an increased risk to develop incident dementia per standard deviation faster rate of decline in hippocampal volume [left hippocampus 1.6 (95% confidence interval 1.2-2.3, right hippocampus 1.6 (95% confidence interval 1.2-2.1)]. Furthermore, decline in hippocampal volume predicted onset of clinical dementia when corrected for baseline hippocampal volume. In people who remained free of dementia during the whole follow-up period, we found that decline in hippocampal volume paralleled, and preceded, specific decline in delayed word recall. No associations were found in this sample between the rate of hippocampal atrophy, Mini Mental State Examination and tests of executive function. The results suggest that rate of hippocampal atrophy is an early marker of incipient memory decline and dementia. It could

therefore be of additional value as a surrogate biomarker of dementia compared with a single hippocampal volume measurement.

In **Chapter 5** potential risk factors for decline of hippocampal volume on MR images were studied in the Rotterdam Scan Study. The same group of 518 persons was used as in Chapter 4. Persons carrying the APOE $\varepsilon 4$ allele had lower hippocampal volumes than persons with the $\varepsilon 3\varepsilon 3$ genotype, but the rate of decline was not influenced by APOE genotype. In persons who did not use antihypertensive treatment, both a high (>90 mmHg) or a low (<70 mmHg) diastolic blood pressure was associated with a faster decline in hippocampal volume. Also, white matter lesions (WMLs) on baseline MRI were associated with a faster decline in hippocampal volume. Vascular factors were associated with rate of hippocampal atrophy, suggesting that accumulating Alzheimer pathology in the hippocampus could be partly due to vascular risk factors.

Chapter 6 dealt with the automated measurement of local white matter lesion volume. It has been hypothesized that WMLs at different locations may have different etiology and clinical consequences. Several approaches for the measurement of local WML load have been proposed in the literature, most of which rely on a distinction between lesions in a periventricular region close to the ventricles and in a subcortical zone further away.

A novel automated method was presented for regional WML volume quantification in MR images. The method segments and measures the white matter lesion volume in 43 regions defined by location and distance to the ventricles, which allows a more detailed study of local lesion load. The potential of the method was demonstrated by analyzing the effect of blood pressure on the regional lesion WML load in 490 elderly subjects taken from the Rotterdam Scan Study. The method was also compared to two commonly-used techniques to assess the periventricular and subcortical lesion load. The main finding was that high diastolic blood pressure was mainly associated with lesion load in the vascular watershed area that forms the border between the periventricular and subcortical regions. This is compatible to the associations between blood pressure and periventricular/subcortical load computed for the same data, and corresponds with findings reported in the literature. However, the proposed method allows analyses with a higher spatial resolution.

7.2 Discussion of contributions

In the introduction, the four most important contributions of this thesis were presented. This section revisits these, and discusses how they were achieved in the work summarized in the previous section.

7.2.1 Methodological extension of atlas-based segmentation

The first contribution is the introduction of three extensions to atlas-based segmentation, which increase the accuracy and robustness of this paradigm.

Global optimization of regularized atlas-based models

In Chapter 2 a segmentation method was presented based on multi-atlas registration and a statistical intensity model implemented with a voxel classifier. The components were combined in a regularized probabilistic framework that was optimized with graph cuts. Previously published brain segmentation methods that used similar models relied on optimizers like iterative conditional modes [Chupin et al., 2009, Fischl et al., 2002] or mean field approximation [Van Leemput et al., 1999]. The main advantage of graph cuts over these techniques is that it is guaranteed to find an optimal solution. Furthermore, it is very easy to implement and a C++ version is publicly available. Although graph cuts have been widely applied in the field of computer vision [Boykov and Kolmogorov, 2004], they were relatively little used in medical image processing until recently. The paper that was the source of Chapter 2, was one of the first journal publications to present the application of graph cuts to brain segmentation. Also the other segmentation methods introduced and applied in this thesis use graph cuts to find an optimal solution.

Validation experiments showed that the atlas&intensity- and atlas&appearance-based methods of Chapters 2 and 3, which both use graph cuts, produce accurate and robust results with respect to manual segmentations. Furthermore, due to the addition of the intensity and appearance models, as well as the regularization, they outperformed methods based on atlas registration alone. However, the influence of the regularization was not directly quantified in these evaluations. A comparison with multi-atlas-based segmentation showed that the accuracy of the two regularized methods is superior, but also suggests that their spatial model (which is based on multi-atlas registration) has the largest influence on the overall segmentation quality (see Sections 2.3.4 and 3.3.4). Visual inspection showed that the regularization did lead to smoother segmentations. Moreover, this was done at little extra computational effort compared to the methods' other components.

Atlas&appearance-based segmentation

In Chapter 3 an atlas&appearance-based segmentation technique was presented and validated. This method combines multi-atlas registration and a voxel classifier based on high-dimensional local appearance features, instead of the intensity-based classifier that was used in the atlas&intensity-based method of Chapter 2. The main motivation for replacing the intensity model with this appearance model was to expand the number of brain structures that can be segmented, and to increase the method's robustness to registration errors.

Validation experiments showed that the method not only accurately segments the hippocampus, but also the cerebellum, a structure with more complex intensity patterns that cannot be properly modeled by the atlas&intensity-based method (see Section 3.3.4). This provides compelling evidence that a combined atlas and appearance model is suitable to perform a whole-brain segmentation. The appearance model was also shown to be better capable of handling large registration errors (See Figure 3.6).

However, the increased robustness did not result in a large improvement of the evaluation measures. Large registration errors occurred in about 10% of the segmented hippocampi, and therefore did not have a large influence on the mean segmentation quality (see Section 3.3.4). Therefore, if the structure of interest has a simple intensity distribution that has little overlap with the background, the atlas &intensity-based method performs on average almost as good as the atlas& appearance-based method, and is faster to compute.

Atlas&intensity-based measurement of hippocampal decline

The last methodological extension presented in this thesis was an adaptation of the atlas&intensity-based method of Chapter 2 to segment brain structures in sequential MR images. The baseline segmentation was performed by combining a spatial probability map created with multi-atlas registration, a statistical intensity model, and a regularizer. The baseline image was then non-rigidly registered to all follow-up scans, and the resulting deformations were used to warp the baseline spatial probability map. The final segmentations were then obtained by combining the warped baseline maps with the same statistical intensity model and regularizer.

Most existing methods for measuring the brain structure atrophy rate are based only on non-rigid registration [Barnes et al., 2008]. The major advantage of the sequential atlas&intensity method is that its intensity model can correct for registration errors. Unfortunately, no manually segmented sequential images were available, and therefore no direct evaluation of the segmentation accuracy could be performed. However, this thesis does provide evidence that the method is suitable for the analysis of longitudinal neuroimaging studies.

First of all, the experiments in Section 2.3.4 show that an intensity model increases the segmentation accuracy in a single image. There is little reason to believe that this is not the case in sequential images, although it is not possible to estimate how large this improvement will be. Furthermore, the rates of hippocampal atrophy measured by the sequential atlas&intensity-based method in Chapters 4 and 5 are in line with previous studies that relied on manual segmentation [Barnes et al., 2009]. Finally, several associations were computed that gave plausible results. These will be discussed in more detail in 7.2.3.

The method did have problems with follow-up images that were acquired with a different MR sequence than the baseline image. An experiment with a small set of subjects scanned with both sequences within a short period of time showed that it overestimated the rate of atrophy (see Section 4.2.3). Fortunately, the relative rates of atrophy within a population were maintained, so the measurements could still be used as Z-scores. However, it cannot be assumed that this will be the case for different datasets. Additional experiments would be needed to ascertain whether the method is really usable to measure decline in images acquired with different sequences or scanners.

7.2.2 Automated versus manual hippocampus segmentation in neuroimaging studies

Chapter 2 also evaluated whether automated brain segmentation can replace manual segmentation. The atlas&intensity-based method was applied to a Rotterdam Scan Study cohort that was previously used to study the relation between manually measured hippocampus volume, cognitive decline, and several risk factors. Tables 2.2 and 2.3 show that the automated method reproduced most of the associations, including those involving the APOE ε4 allele, memory performance and the risk of developing dementia.

The evaluation of the accuracy of the method's volume measurement on 20 manually segmented images also provides evidence of its performance in neuroimaging studis (Section 2.3.4). The results of this experiment revealed a tendency to underestimate large volumes and overestimate small volumes. This bias was also seen for the atlas&appearance-based method of Chapter 3. As it results in a reduction of the sensitivity, it could very well explain the failure to replicate some of the associations in Tables 2.2 and 2.3. Taken together, these results show that the atlas&intensity-based method can be used to replace manual segmentation to study associations of brain structure volume. However, it might not be sensitive enough to find more subtle relations.

Another caveat is that the intensity model of the atlas&intensity-based method requires training data with intensity characteristics comparable to the target images that need to be segmented. The same is true for the appearance model of Chapter 3. Therefore, these methods can only be applied to neuroimaging studies if a subset of manually segmented images is available, and the data is acquired with the same scanner and MR sequence.

7.2.3 Analysis of the rate of hippocampal atrophy

The sequential atlas&intensity-based segmentation method allowed the measurement of hippocampal atrophy rates in the Rotterdam Scan Study. Since this task involved 947 images, manual segmentation would have been a very time-consuming, laborious process. The measurements were used to analyze risk factors of hippocampal decline (Chapter 5), as well as its relation to cognitive decline followed over a ten-year period (Chapter 4). Previous studies on these associations were of limited size and not performed in a general population setting.

From these studies it was known that the rate of atrophy is higher in persons who were diagnosed with AD. However, in Chapter 4 it was demonstrated that increased atrophy also precedes clinical symptoms. The longitudinal population-based study design also enabled us to show that a decline in hippocampal volume paralleled, or could even precede cognitive decline in the subjects who remained free of dementia.

Previous work on risk factors of hippocampal decline was mainly focused on the APOE genotype. Chapter 5 studied not only the influence of APOE on the rate of hippocampal atrophy, but also the influence of vascular factors. The results

showed that hypertension, which had been associated previously with white matter lesions and a low hippocampal volume, is also related to an increased rate of hippocampal atrophy.

7.2.4 Automated measurement of regional white matter lesion volume

Automated brain structure segmentation as proposed in this thesis can also be applied to investigate local WML volume. For example, the often-used distinction between periventricular and subcortical lesions requires a segmentation of the ventricles. In Chapter 6, the atlas&intensity-based method of Chapter 2 and a multi-atlas registration were used to perform a measurement of WML volume in 43 regions, based on the distance and location with respect to the ventricles (Figure 6.1).

The potential of the method was demonstrated by analyzing the relation between blood pressure and regional WML volume. The smaller regions allowed a more detailed perspective on the effect of this risk factor on WMLs at different locations in the brain. For example, the results showed an association between diastolic blood pressure and lesion load in the vascular watershed area around the ventricles, which forms the boundary between the periventricular and subcortical zones (Figure 6.4). This conclusion could not have been drawn if the lesion load was measured in the periventricular and subcortical zones only. Automated brain structure segmentation is therefore not only a tool to automate local WML analysis, but can also increase its spatial resolution.

7.3 Conclusions and future research

The work presented in this thesis shows that a combination of multi-atlas registration, a statistical intensity model, and regularization can produce accurate and robust brain structure segmentations. These atlas&intensity-based techniques outperform multi-atlas-based methods with little additional computation costs. This type of method is accurate enough to replace manual segmentation in volumetry studies, or to define regions of interest for the classification of WMLs. Furthermore, the combination of atlas registration and an intensity model can also be applied to measure the rate of brain structure atrophy from sequential MR images. The intensity model used in atlas&intensity-based segmentation cannot model brain structures with similar intensity distributions as the background. Replacing the intensity model with an appearance model implemented with a high-dimensional voxel classifier can overcome this limitation.

The methods presented in this thesis have two shortcomings: they require training images with intensities that are similar to the unlabeled target images, and they yield a binary segmentation. Overcoming these limitations provides a promising direction for future work.

As noted in Section 1.2, segmentation methods that depend on a specific training set are unlikely to be used outside the institutions where they were developed.

Furthermore, it also reduces their ability to process data from daily clinical routine, previous research, or multi-center studies.

This problem could be addressed by a normalization procedure that remaps the intensities of the unlabeled target image to match the training data. The two publicly available brain segmentation methods FIRST and Freesurfer both use this method to handle images with a wide range of intensity characteristics [Fischl et al., 2004, Han and Fischl, 2007, Patenaude, 2007]. An alternative is to use atlas-based segmentation to sample intensity or feature values from the target image and build a target-specific intensity or appearance model. If a similarity measure like normalized mutual information is used to drive the registration, this procedure will be relatively robust against intensity differences between the atlas and the target image. This strategy is very common in brain tissue segmentation methods [Cocosco et al., 2003, Van Leemput et al., 1999, Vrooman et al., 2007, Zhang et al., 2001], but has not yet been used much for brain structure segmentation [Wolz et al., 2010].

Multi-structure models that segment the entire brain would also increase the applicability of automated methods. However, the methods presented in this thesis cannot be used to label multiple structures as the graph cuts optimizer can only solve binary segmentation problems. Recently, alternative combinatorial optimizers have been introduced in computer vision like FastPD [Komodakis and Tziritas, 2007]. Although these techniques are not guaranteed to find a global optimum, they have been shown to be more robust than iterative conditional modes. Therefore, they are an excellent basis for a whole-brain segmentation method.

References

J. Barnes, J. Foster, R. G. Boyes, T. Pepple, E. K. Moore, J. M. Schott, C. Frost, R. I. Scahill, and N. C. Fox. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage*, 40(4):1655–1671, 2008.

J. Barnes, J. W. Bartlett, L. A. van de Pol, C. T. Loy, R. I. Scahill, C. Frost, P. Thompson, and N. C. Fox. A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. *Neurobiol Aging*, 30(11):1711–1723, 2009.

Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans Pattern Anal Mach Intell*, 26(9):1124–1137, 2004.

M. Chupin, A. Hammers, R. S. N. Liu, O. Colliot, J. Burdett, E. Bardinet, J. S. Duncan, L. Garnero, and L. Lemieux. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *NeuroImage*, 46(3):749–761, 2009.

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

B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

B. Fischl, D. H. Salat, A. J. W. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23 Suppl 1:S69–S84, 2004.

X. Han and B. Fischl. Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Trans. Med. Imaging*, 26(4):479–486, 2007.

N. Komodakis and G. Tziritas. Approximate labeling via graph cuts based on linear programming. *IEEE Trans Pattern Anal Mach Intell*, 29(8):1436–1453, 2007.

B. Patenaude. *Bayesian Statistical Models of Shape and Appearance for Subcortical Brain Segmentation*. PhD thesis, University of Oxford, 2007.

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

H. A. Vrooman, C. A. Cocosco, F. van der Lijn, R. Stokking, M. A. Ikram, M. W. Vernooij, M. M. B. Breteler, and W. J. Niessen. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *NeuroImage*, 37(1):71–81, 2007.

R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, and A. D. N. Initiative. LEAP: learning embeddings for atlas propagation. *NeuroImage*, 49(2):1316–1325, 2010.

Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging*, 20(1):45–57, 2001.

Samenvatting

Het doel van dit proefschrift was de ontwikkeling van nauwkeurige en robuuste methodes voor de segmentatie van hersenstructuren in MRI scans en hun validatie en toepassing in de context van een grote neuroimaging studie. In de voorgaande hoofdstukken werden drie nieuwe atlas-gebaseerde methodes geïntroduceerd voor de segmentatie van enkele of sequentiële beelden. De methodes zijn toegepast op data van de Rotterdam Scan Study om associaties te bestuderen van hippocampaal volume en volume verlies, en om de lokale wittestoflaesielast te kwantificeren.

Hoofdstuk 2 introduceert een segmentatiemethode op basis van de minimalisatie van een energiefunctionaal met een intensiteits- en een a prioriterm die zijn afgeleid van manueel gelabelde trainingsbeelden. De intensiteitsenergie wordt geïmplementeerd met een voxel classifier die de intensiteitsdistributies van de voorgrond- en achtergrondklasse leert uit de trainingsdata. De spatiële a prioriterm wordt verkregen door de trainingsbeelden naar het ongelabelde doelbeeld te registreren, de trainingslabels te vervormen en te middelen. De a prioriterm heeft ook een regularisatiecomponent die gladdere segmentaties bevordert. De energiefunctionaal die is opgebouwd uit deze termen kan globaal worden geminimaliseerd met behulp van graph cuts.

De methode werd geëvalueerd met data afkomstig van de Rotterdam Scan Study, een grootschalige neuro-imaging studie naar ouderdomsziekten. Uit dit onderzoek werden twee datasets geselecteerd: een kleine set van 20 MRI-beelden waarin twee experts de hippocampi hadden gesegmenteerd, en een grote set van 498 beelden waarvoor manueel gemeten hippocampusvolumes beschikbaar waren, maar geen segmentaties. Met deze set waren in het verleden significante verbanden aangetoond tussen hippocampaal volume, cognitief functioneren en de ontwikkeling van dementie. Kruisvalidatie experimenten met de gelabelde set lieten een Dice overlap zien van 0.852 en 0.864, en een gemiddelde afstand sfout van 0.40 en 0.36 mm voor respectievelijk de linker- en rechterzijde. 83% van de segmentatie van de grote set werd als goed beoordeeld door een expert. Verder werden de automatische gemeten volumes van deze set gebruikt om een aantal van de manuele associatiestudies te herhalen. De automatisch gemeten volumes lieten vergelijkbare associaties zien met cognitieve functies en de ontwikkeling van dementie. Tenslotte presteerde de atlas&intensiteits-gebaseerde methode beter dan een atlasgebaseerde methode in een vergelijkingsexperiment.

Hoofdstuk 3 beschrijft een segmentatiemethode voor hersenstructuren gebaseerd op informatie over hun locatie en de verschijning op MRI-beelden. Het spatiële model wordt gegenereerd door registratie en middeling van meerdere

manueel gesegmenteerde atlassen. De verschijning van de hersenstructuur wordt gemodelleerd met een classifier met Gaussische scale-space features. De twee componenten worden gecombineerd met een regularisatie term in een Bayesiaans model dat globaal kan worden geoptimaliseerd met graph cuts. Met het verschijningsmodel kan de methode structuren segmenteren met complexe intensiteitsdistributies en is hij minder kwetsbaar voor fouten in het spatiële model.

De methode werd getest door de hippocampus en het cerebellum te segmenteren in kruisvalidatie-experimenten met twee manueel gelabelde datasets van verschillende MRI scanners. Daarnaast werd de methode op dezelfde data vergeleken met drie alternatieven gebaseerd op atlas registratie, atlas registratie plus een intensiteitsmodel en Freesurfer. De atlas&verschijnings-gebaseerde methode produceert nauwkeurige resultaten met een Dice overlap van 0.95 voor het cerebellum en 0.87 voor de hippocampus. Dit was significant beter dan de atlas-gebaseerde methode en vergelijkbaar met de atlas&intensiteitscombinatie. Vergeleken met de laatste techniek is de atlas&verschijnings-gebaseerde methode breder inzetbaar en robuuster tegen fouten in het spatiele model. De vergelijking met Freesurfer was gebaseerd op correlaties van de volumemetingen. Deze waren vergelijkbaar voor beide methodes.

MRI scans van Alzheimerpatiënten en personen met mild cognitive impairment (MCI) tonen vaak een geatrofieerde hippocampus. Zelfs bij ouderen zonder symptomen van dementie is een kleinere hippocampus een risicofactor voor het ontwikkelen van de ziekte van Alzheimer. Aan de andere kant krijgt ook niet iedereen met een kleine hippocampus dementie. Eerder onderzoek heeft laten zien dat hippocampi van Alzheimerpatiënten sneller atrofiëren dan die van personen zowel zonder klachten als met MCI. **Hoofdstukken 4 en 5** onderzoeken de potentie van volumeafnamemetingen in twee of meerdere sequentiële MRI scans als biomarker voor het ziekteproces.

Het is nog onduidelijk of een snellere volumeafname van de hippocampus ook gerelateerd is aan het later ontwikkelen van cognitieve achteruitgang of de ziekte van Alzheimer. **Hoofdstuk 4** onderzoekt deze associaties bij 518 deelnemers van de Rotterdam Scan Study (60-90 jaar, 50%vrouw). Van deze groep werden scans gemaakt in 1995-1996, 1999-2000 (244 deelnemers) en in 2006 (185 deelnemers). Met behulp van een automatische methode gebaseerd op atlas- en intensiteitsinformatie werden de hippocampusvolumes gemeten in al deze sequentiële scans. Geen van de deelnemers leed aan dementie ten tijde van de eerste scan. Gedurende tien jaar werden vier neuropsychologische tests afgenomen en werd de groep gescreend op de ontwikkeling van dementie.

In de follow-up periode van 4360 persoonsjaren (gemiddeld 8.4, bereik 0.1-11.3) ontwikkelden 50 deelnemers een vorm van dementie (waarvan 36 de ziekte van Alzheimer). We vonden een verhoogd risico voor dementie per standaard deviatie grotere volumeafname (links: 1.6 (95% betrouwbaarheidsinterval 1.2;2.3), rechts: 1.6 (95% betrouwbaarheidsinterval 1.2;2.1)). Ook na correctie voor het aanvangsvolume bleef volumeafname voorspellend. In de deelnemers zonder dementie

ging een snellere volumeafname gepaard en vooraf aan een snellere achteruitgang van geheugen. Dit verband werd niet gevonden met de scoreafnames van de Mini Mental State Examination of executieve functietests. De resultaten suggereren dat de versnelde afname van hippocampusvolume een marker is van beginnende geheugenklachten en dementie. Daarmee hebben meerdere metingen toegevoegde waarde als biomarker voor dementie vergeleken met een enkele volumemeting.

In **hoofdstuk 5** werden een aantal risicofactoren onderzocht van hippocampaal volumeverlies zoals gemeten op sequentiële MRI scans van de Rotterdam Scan Study. Dezelfde groep van 518 deelnemers werden gebruikt als in hoofdstuk 4. Dragers van een APOE $\varepsilon 4$ allele hadden kleinere hippocampi dan dragers van het APOE $\varepsilon 3\varepsilon 3$ genotype, maar dit had geen invloed op de snelheid waarmee het volume van de structuur afnam. Bij personen die geen bloeddrukverlagende middelen slikten was zowel een hoge (>90 mmHg) als een lage (<70 mmHg) diastolische bloeddruk geassocieerd met en groter volumeverlies. Daarnaast was ook de hoeveelheid wittestof laesies (WSLs) op de eerste MRI gerelateerd aan een grotere volumeafname. Deze associaties tussen vasculaire risicofactoren en het volumeverlies suggereren dat deze factoren bijdragen aan de progressie van Alzheimer pathologie in de hippocampus.

Hoofdstuk 6 behandelt de automatische meting van lokale WSL volumes. Een aantal eerdere artikelen hebben gesuggereerd dat WSLs op verschillende locaties een andere etiologie en klinische gevolgen hebben. Om deze hypothese verder te onderzoeken zijn er verschillende methodes gepubliceerd voor het meten van de lokale WSL-last. Voor het overgrote deel zijn deze gebaseerd op een onderscheid tussen laesies in een periventriculaire zone grenzend aan de ventrikels en in een subcorticale zone op grotere afstand.

In dit werk wordt een methode gepresenteerd om de WSL-last per regio automatisch te kwantificeren. De methode segmenteert en meet het aantal laesies in 43 regio's op basis van hun afstand en oriëntatie ten opzichte van de ventrikels. Hierdoor is een gedetailleerdere analyse mogelijk van de locatie waar laesies worden gevonden. De potentie van de methode wordt gedemonstreerd door een analyse van de associatie tussen bloeddruk en regionaal WSL volume in 490 deelnemers van de Rotterdam Scan Study. De methode werd ook vergeleken met twee veelgebruikte technieken om de periventriculaire en subcorticale WSL last te kwantificeren. De belangrijkste bevinding is dat hogere diastolische bloeddruk voornamelijk is geassocieerd met laesies in de vasculaire waterscheiding die de periventriculaire en subcorticale zones scheidt. Dit resultaat is in overeenstemming met de relaties tussen bloeddruk en de subcorticale/periventriculaire last die werden gevonden met de andere twee methodes en met eerdere studies uit de literatuur. Daarnaast heeft de gepresenteerde techniek een hogere spatiële resolutie dan de meeste alternatieven.

Dankwoord

Op zoek naar een promotieplek kwam ik er al snel achter dat ik een kandidaat "met een verhaal" was. Door mijn uitstap naar de journalistiek en wetenschapsvoorlichting had ik al twee jaar geen onderzoek meer gedaan. Vandaar dat ik in de eerste plaats mijn promotor Wiro Niessen wil bedanken voor zijn vertrouwen. Wiro, ik waardeer het heel erg dat je open stond voor dat verhaal en bedankt voor alle raad en de gouden tip om eens naar die graph cuts te gaan kijken. Jouw enthousiasme en humor zijn aanstekelijk; de explosieve groei van de Biomedical Imaging Group Rotterdam lijkt me daarvan het beste bewijs.

Mijn co-promotor Henri Vrooman wil ik graag bedanken voor zijn hulp tijdens het hele traject. Henri, zonder jou zou ik direct al zijn verzopen in alle data en tools die nodig zijn voor beeldanalyse van een grote neuroimaging studie. En dat brengt me bij Monique Breteler: Monique, dankzij jou heb ik in mijn onderzoek kunnen werken met één van de mooiste datasets ter wereld; het WK voetbal is toch net even iets spannender dan de zaterdagamateurs. Daarnaast heb ik veel van je geleerd over epidemiologie, wat ik een (bijna) net zo'n mooi vak ben gaan vinden als beeldverwerking.

Mijn dank gaat ook uit naar de leden van mijn leescommissie: Monique, dr. Boudewijn Lelieveldt en prof. Peter Koudstaal. Dank dat jullie mijn werk hebben willen lezen. Prof. Miriam Sturkenboom, prof. Gabriel Krestin en prof. Mark van Buchem wil ik graag bedanken voor het plaatsnemen in mijn promotiecommissie. Finally, prof. Daniel Rueckert, I am honored that you were willing to come to Rotterdam for my defence. Also, thank you for your help with your registration method; that really got me hooked on atlas-based segmentation. Alexander Hammers was also very important in this respect. Thank you for sending us your magnificent atlas and for your encouraging comments on my work.

Marleen en Stefan: ik ben stikjaloers op het gemak waarmee jullie kunnen praten en denken over registratie en patroonherkenning, maar ondertussen profiteer ik er mooi wel van. Jullie komst heeft mijn onderzoek naar een hoger plan getild en daarnaast zijn jullie allebei ook nog eens de goedheid zelve. Michiel: co-auteurs zijn we pas op het allerlaatst geworden, maar ik heb ontzettend veel inspiratie, kennis en plezier van onze discussies en gesprekken gehad; ik ben dan ook heel blij dat je mijn paranimf wil zijn. Neem jij de bul mee als Elkes vliezen breken tijdens de verdediging?

Dat ik mijn promotieonderzoek met zoveel plezier heb gedaan, heb ik zeker ook te danken aan mijn twee meest langdurige kamergenoten Renske en Marius. Dank jullie wel voor de gezelligheid in en buiten Ee2167! Maar ook die momenten

dat het even minder ging in en buiten Ee2167 heb ik onwijs veel aan jullie gehad. Hakim, leuk dat je de neurogroep komt versterken en dat je verder gaat met mijn segmentaties.

Aad van der Lugt, dank voor je waardevolle commentaar op mijn manuscripten, en voor je karakteristieke bezoekjes aan onze kamer op weg van of naar Wiro. Desiree de Jong, jou wil ik graag bedanken voor je hulp bij het regelwerk voor het formele gedeelte van deze promotie. Maaike en Joyce: jullie waren de eersten die mijn methodes hebben getest op data uit andere studies. Het was me een waar genoegen om met jullie samen te werken. Eva, jammer genoeg hebben we nooit de grote classificatie-registratie-neuro-long fusie van de grond gekregen, maar het was supergezellig om bij te kunnen kletsen als "schaduw-AIOs". To all other people from BIGR, Medical Informatics and Radiology: there are just too many of you, but thank you all for the great atmosphere!

De samenwerking met de studenten en postdocs van epidemiologie was voor mij één van de leukste kanten van dit onderzoek. Allereerst kan ik natuurlijk niet om mijn partner in hippocrime heen: Tom, dank je wel voor je onverwoestbare optimisme. Ik hoop dat we de komende tijd door kunnen gaan met oogsten wat we hebben gezaaid de afgelopen jaren. Meike en Arfan: van dat vooroordeel over arrogante artsen heb ik nooit iets gemerkt bij jullie: jullie belangstelling voor het werk van ons beeldverwerkers heb ik altijd fantastisch gevonden. Kom maar weer snel terug naar de 21e, ik mis de legendarisch Arfan swagger, de verbeten tweevingerige Meike type-stijl en de koffie. Jory, bedankt voor de gesprekken en alle moeite die je in de manuele segmentaties hebt gestoken. Ben Verhaaren: dank voor je werk aan het laatste hoofdstuk van dit proefschrift.

Manfred Kayser gave me the opportunity to finish this thesis while I worked in his forensics lab at Erasmus MC. Manfred, I really admire your broad perspective on research. I do not think there are a lot of other forensic biologists that would hire an image processing guy like me. On top of that you have provided me with a new set of great colleagues.

Anna, Maaike, Ruben, Lotte, Bas, Bart, Els, Erik, Susan, de Runderfunkers, Gilbert, Willemijn, Tanja, Oscar, Sascha, Josine, Anne Marije: bedankt voor jullie vriendschap, scherpe opmerkingen, muziek, lekkere eten, mooie foto's en aanverwante artikelen. Meneer Taco, zonder jou waren er geen Rundfunk djs en had ik een goede vriend minder. Sono, wat goed dat je mijn paranimf wil zijn! Jouw vrolijkheid is de beste remedie tegen de promotiezenuwen.

Pap, Mam, bij jullie kan ik altijd thuiskomen: dank je wel voor alles. Hidde en Jikke, we zien elkaar misschien niet zo vaak, maar toch zijn jullie mijn favoriete gezwister. Ik ken geen grappigere mensen dan jullie. Jaap, Els, Jelle, Bart, Lucie, Joost, Pam: ik heb me van het begin af aan welkom gevoeld bij jullie!

En dan natuurlijk Elke: juni 2005 begon met de eerste experimenten van mijn promotie en een afspraakje op de Grote Markt. Eind oktober 2010 heb ik m'n bul en deel ik een huis met de twee mooiste meisjes. Alles, alles, jij hebt het gedaan!

Fedde van der Lijn, Den Haag, 22 Juli 2010.

Publications

Papers in international journals

- T. den Heijer*, F. van der Lijn*, P.J. Koudstaal, A. Hofman, A. van der Lugt, G.P. Krestin, W.J. Niessen, and M.M.B. Breteler. A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain*, 133(4):1163–1172, 2010. *joint first authorship

- F. van der Lijn, T. den Heijer, M.M.B. Breteler and W.J. Niessen. Postmenopausal hormone therapy and regional brain volumes: the WHIMS-MRI study. *Neurology*, 73(18):1514, 2009.

- R. de Boer, H. Vrooman, F. van der Lijn, M. W. Vernooij, M. A. Ikram, A. van der Lugt, M. M. B. Breteler and W.J. Niessen. White matter lesion extension to automatic brain tissue segmentation on MRI. *NeuroImage*, 45(4):1151–61, 2009.

- F. van der Lijn, T. den Heijer, M.M.B. Breteler and W.J. Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage*, 43(4):708–720, 2008.

- M.A. Ikram, H. Vrooman, M.W. Vernooij, F. van der Lijn, A. Hofman, A. van der Lugt, W.J. Niessen and M.M.B. Breteler. Brain tissue volumes in the general elderly population: The Rotterdam Scan Study, *Neurobiology of Aging*, 29(6):882-90, 2008.

- H. Vrooman, C.A. Cocosco, F. van der Lijn, R. Stokking, M.A. Ikram, M.W. Vernooij, M.M.B. Breteler and W.J. Niessen. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification, *NeuroImage*, 29(6):882-90, 2007.

- F. van der Lijn, M. de Bruijne, S. Klein, T. den Heijer, Y.Y. Hoogendam, A. van der Lugt, M.M.B. Breteler and W.J. Niessen. Automated brain structure segmentation based on atlas registration and appearance models. *Submitted*.

- T. den Heijer*, F. van der Lijn*, M.A. Ikram, P.J. Koudstaal, A. van der Lugt, G.P. Krestin, H.A. Vrooman, A. Hofman, W.J. Niessen, and M.M.B. Breteler. Baseline predictors of hippocampus decline on MRI over a 10-year follow-up. *Submitted*. *joint first authorship

- F. van der Lijn, B.F. Verhaaren, M.A. Ikram, S. Klein, H.A. Vrooman, M.W. Vernooij, A. Hammers, D. Rueckert, A. van der Lugt, M.M.B. Breteler and W. Niessen. Automated measurement of local white matter lesion volume. *Submitted*.

Papers in conference proceedings

- R. de Boer, M. Schaap, F. van der Lijn, H. Vrooman, M. de Groot, M.W. Vernooij, M.A. Ikram, E.F.S. van Velsen, A. van der Lugt, M.M.B. Breteler and W.J. Niessen. Statistical Analysis of Structural Brain Connectivity. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention 2010*, in press.
- S. Klein, M. Loog, F. van der Lijn, T. den Heijer, A. Hammers, M. de Bruijne, A. van der Lugt, R.P.W. Duin, M.M.B. Breteler and W.J. Niessen. Early diagnosis of dementia based on intersubject whole-brain dissimilarities. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pp 249–252, 2010.
- F. van der Lijn, M. de Bruijne, Y.Y. Hoogendam, S. Klein, K. Hameeteman, M.M.B. Breteler and W.J. Niessen. Cerebellum Segmentation in MRI Using Atlas Registration and Local Multi-Scale Image Descriptors. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pp 221–224, 2009.
- F. van der Lijn, T. den Heijer, M.M.B. Breteler and W.J. Niessen. Combining graph cuts, atlas registration, and voxel classification for hippocampus segmentation in MR images. In *Mathematical Methods of Biomedical Image Analysis*, 2007.
- F. van der Lijn, M.W. Vernooij, M.A. Ikram, H.A. Vrooman, D. Rueckert, A. Hammers, M.M.B. Breteler and W.J. Niessen. Automated localization of periventricular and subcortical white matter lesions. In *Proceedings of SPIE Medical Imaging*, 2007.
- R. de Boer, F. van der Lijn, H. Vrooman, M.W. Vernooij, M.A. Ikram, M.M.B. Breteler and W.J. Niessen. Automatic segmentation of brain tissue and white matter lesions in MRI. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pp 652–655, 2007.
- F. van der Lijn, F. Roth, and M. Verhaegen. Estimating the impulse response of buried objects from ground penetrating radar signals, In *Proceedings of SPIE Detection and Remediation Technologies for Mines and Minelike Targets*, pp. 387-394, 2003.

Abstracts in conference proceedings

- J.M.G. Florisson, F. van der Lijn, H. Vrooman, W.J. Niessen, I.M.J. Mathijssen and M.H. Lequin. Cerebellum Volume in Children with Syndromic Cran-

iosynostosis. In *Annual Meeting of the European Society of Paediatric Radiology*, 2010.

- J.M.G. Florisson, F. van der Lijn, H. Vrooman, W.J. Niessen, M.H. Lequin and I.M.J. Mathijssen. Brain Malformations in Children with Syndromic Craniosynostosis. In *Annual Meeting of the European Society of Paediatric Radiology*, 2010.

- B.F.J. Verhaaren, F. van der Lijn, S. Klein, A. van der Lugt, W.J. Niessen and M.M.B. Breteler. Automated classification of periventricular and subcortical white matter lesions on MRI. In *Alzheimer's Association International Conference on Alzheimer's Disease*, 2009.

- T. den Heijer, F. van der Lijn, P.J. Koudstaal, A. Hofman, W.J. Niessen and M.M.B. Breteler. A ten year follow-up of hippocampal volume on MRI in early dementia and cognitive decline. In *Alzheimer's Association International Conference on Alzheimer's Disease*, 2009.

- R. de Boer, H. Vrooman, F. van der Lijn, M.W. Vernooij, M.A. Ikram, M.M.B. Breteler and W.J. Niessen. Automatic multiple-atlas-based segmentation of brain tissue and white matter lesions in MRI. In *Second Dutch Conference of Biomedical Engineering*, 2009.

- F. van der Lijn, M.W. Vernooij, M.A. Ikram, H.A. Vrooman, D. Rueckert, A. Hammers, M.M.B. Breteler and W.J. Niessen. Automated localization of periventricular and subcortical white matter lesions. In *Second Dutch Conference of Biomedical Engineering*, 2007.

- H.A. Vrooman, C.A. Cocosco, F. van der Lijn, R. Stokking, M.A. Ikram, M.W. Vernooij, M.M.B. Breteler and W.J. Niessen. Multi-spectral MRI brain tissue segmentation: Training on manually labeled scans versus automatic atlas-based training on single subject. In *First Dutch Conference on Bio-Medical Engineering*, 2007

- R. de Boer, H. Vrooman, F. van der Lijn, M.W. Vernooij, M.A. Ikram, M.M.B. Breteler and W.J. Niessen. Automatic multiple-atlas-based segmentation of brain tissue and white matter lesions in MRI. In *Scientific Assembly and Annual Meeting of the Radiological Society of North America*, 2007.

- M.A. Ikram, H.A. Vrooman, M.W. Vernooij, F. van der Lijn, W.J. Niessen and M.M.B. Breteler. Quantitative measures of brain tissue volumes in the elderly: The Rotterdam Scan Study. In *Alzheimer's Association International Conference on Alzheimer's Disease*, 2006.

PhD Portfolio

PhD period: 2005 - 2010
Departments: Radiology & Medical Informatics
Graduate school: ASCI

	Year	ECTS
In-depth courses		
Principles of Research in Medicine and Epidemiology, NIHES, Rotterdam	2005	0.7
Introduction to Data-analysis, NIHES, Rotterdam	2005	0.7
Regression Analysis, NIHES, Rotterdam	2005	1.4
Knowledge Driven Image Segmentation, ASCI, Leiden	2005	4
Non-Linear Shape Modelling, ITU, Copenhagen,	2005	4
Introduction to MevisLab, MeVis, Bremen	2006	0.5
Advanced Pattern Recognition, ASCI, Delft	2006	4
Advances in Clinical Neuroepidemiology, NIHES, Rotterdam	2007	0.7
FSL & Freesurfer Course, FMRIB, Cardiff	2007	1
Presentations on international conferences		
IEEE International Symposium on Biomedical Imaging, Boston (oral presentation)	2009	1
IEEE Workshop on Mathematical Methods of Biomedical Image Analysis, Rio de Janeiro (poster)	2007	1
SPIE Medical Imaging, San Diego (poster)	2007	1
International conferences		
IEEE International Symposium on Biomedical Imaging, Rotterdam	2010	1
Alzheimer's Association International Conference on Alzheimer's Disease	2009	1
IEEE International Conference on Computer Vision, Rio de Janeiro	2007	1
Annual Meeting of the Organization for Human Brain Mapping, Florence	2006	1

Seminars and workshops		
Machine Learning for Biomedical Image Analysis, Rotterdam	2010	0.3
Statistical Shape Analysis: Theory, Software, and Applications, Boston	2009	0.3
Discrete Optimization methods in Computer Vision, Rio de Janeiro	2007	0.3
Other		
Reviews for NeuroImage, IEEE Transactions of Image Processing, and IEEE Transactions of Medical Imaging	2008-now	3
Visigen Consortium Meeting, Rotterdam (oral presentation),	2008	0.5
Medical Imaging Symposium for PhD Students, Leiden (oral presentation)	2008	0.5
First Dutch Conference on Bio-Medical Engineering, Egmond aan Zee (oral presentation)	2007	1
Teaching activities		
Supervision MSc Project Balinder Paul	2007-2008	5
Teaching Assistant Introduction to Image Processing for Medical Students	2006-2009	3
Teaching Assistant Knowledge Driven Image Segmentation	2005,2007	1
Total		**38.9**

About the Author

Fedde van der Lijn was born February 5, 1977 in Eindhoven, the Netherlands. He studied applied physics at the University of Twente (Enschede, the Netherlands) from 1995 until 2002. His graduation project was on the deconvolution of ground penetrating radar signals for demining. After obtaining his MSc. degree cum laude Van der Lijn followed a post-graduate course on journalism in 2003, which included an internship at the national newspaper Algemeen Dagblad. This resulted in a job as science communication advisor at CWI, the Dutch national research center for mathematics and computer science in Amsterdam. At CWI van der Lijn's interest in research was rekindled and in 2005 he started working as a PhD student at the Biomedical Imaging Group Rotterdam, Erasmus MC, the Netherlands. The result of this work can found in this thesis. Since 2009 he works at the Erasmus MC Department of Forensic Molecular Biology on the automated extraction of facial phenotypes to study the genetic basis of appearance.