University of Szeged
Department of Applied Informatics

Novel MR Image Analysis Strategies: Applications in Multiple Sclerosis

Doctoral Thesis

by

László G. Nyúl

Thesis advisors:

Jayaram K. Udupa, PhD and **Attila Kuba, PhD**

Szeged
2002

Acknowledgments

I carried out most of the research reported here during my stay at the Medical Image Processing Group (MIPG), Department of Radiology, University of Pennsylvania in Philadelphia, PA. Since I did most of my work as a member of MIPG working in collaboration with other members of the Group on certain projects, several times throughout this thesis (especially in Chapter 6), phrases such as "we", "us", "our" refer to the Group.

I am grateful to my thesis advisor, Jayaram K. Udupa for introducing me to the depths of image segmentation, for his leadership and continuous support during these past 5 years. I also thank him for encouraging me to draft and revise my papers on my own from the very beginning, and for his invaluable comments on my writings. I am also grateful to my chief and supervisor, Attila Kuba for directing my eyes to digital image processing when I was his student and for his continuous support ever since. I thank him for the partial relief from other duties while I was writing this thesis, and for taking the time to read and comment on the manuscript.

I thank Luogang Wei his friendly support during the first year of my stay in Philadelphia and his help in getting acquainted with the original Multiple Sclerosis package written for the 3DVIEWNIX system. I also thank Dewey Odhner his prompt help whenever I had questions about or problems with the 3DVIEWNIX system while prototyping and implementing the algorithms. Thanks are due to Tianhu Lei, Alexandre X. Falcão and Punam K. Saha for the very useful discussions and other help I received from them. I acknowledge the help I received from Attila Tanács in image registration matters. I am indebted to my colleagues, Antal Nagy, Marianna Dudásné Nagy, and Emese Balogh for covering my teaching duties during the time I have been abroad and also for bearing with me while this thesis was being written. I thank the help of all other members of both the Medical Image Processing Group, University of Pennsylvania and the Departments of Informatics, University of Szeged, that allowed me to finish this thesis at last.

I am thankful to Dr. Robert I. Grossman, Dr. David C. Hemmy, and Dr. Richard A. Baum, respectively, for the MRI brain, MRI foot, and MR angiography datasets utilized in this research.

My work was supported by funds from the U.S. Government and the Hungarian Government, in particular, by NIH Grants NS 37172 and AR 46902, a DHHS Grant LMO-3502-MOD3, a grant from the Department of Army DAMD 179717271, a grant from the Hungarian Ministry of Education FKFP 0908, and by the Hungarian Eötvös Scholarship.

Abbreviations

2D two dimensions, two dimensional

3D three dimensions, three dimensional

4D four dimensions, four dimensional

AFFIRMATIVE attenuation of fluid by fast inversion recovery with magnetization transfer imaging with variable echoes

AI ambulation index

ANAWM abnormal normal appearing white matter

ANN artificial neural network

BBB blood brain barrier

BG background

BP brain parenchyma

CI confidence interval

CNS central nervous system

CPU central processing unit

CSE conventional spin-echo

CSF cerebro-spinal fluid

CT computerized tomography, computer tomograph

CV coefficient of variation

DICOM Digital Imaging and Communications in Medicine

EDSS Expanded Disability Status Scale

EM expectation maximization

EMT extrameningeal tissue

FIFO first-in first-out

FLAIR fluid attenuation inversion recovery

fMRI functional magnetic resonance imaging

FNVF false negative volume fraction

FOV field of view

FPVF false positive volume fraction

FSE fast spin-echo

Gd Gadolinium

Gd-DTPA Gadolinium (III) diethyltriaminepentaacetic acid (DTPA), an MRI contrast reagent

GM gray matter

GRASS gradient acquisition in steady state

GRE gradient echo imaging

ICC intracranial cavity

ICM iterated conditional modes

I/O input/output

IOI intensity of interest

kNN k-nearest neighbors

LIFO last-in first-out

LS lesion(s)

LUT look-up table

MI mutual information

ML maximum likelihood

MP-RAGE magnetization prepared rapid gradient echo

MR magnetic resonance

MRA magnetic resonance angiography

MRF Markov random field

MRI magnetic resonance imaging

MS Multiple Sclerosis

MT magnetization transfer

MTC magnetization transfer contrast

MTI magnetization transfer imaging

MTR magnetization transfer ratio

NAWM normal appearing white matter

NEX number of excitations

NMI normalized mean intensity

NMSD normalized mean squared difference

PACS Picture Archiving and Communication System

PC personal computer

PD proton density

PET positron emission tomography

SP primary-progressive

RF radio frequency

ROI region of interest

RMS root-mean-squared

RR relapsing-remitting

SE spin-echo imaging

SP secondary-progressive

SPGR spoiled gradient recalled (spoiled GRASS)

T1 first echo

T1E T1 enhanced (with Gd)

T1LV T1 lesion volume

T1WI T1-weighted imaging

T1WIE T1-weighted imaging with Gadolinium enhancement

T2 second echo

T2LV T2 lesion volume

T2WI T2-weighted imaging

TE time delay between excitation and echo maximum

TLV total visible lesion volume

TPVF true positive volume fraction

TR time to repetition

VOI volume of interest, value of interest

WM white matter

XOR exclusive OR

Contents

1 Introduction **1**

2 Standardizing the MR Image Intensity Scale **5**
2.1 The Problem with MR Image Intensities 5
2.2 The Standardizing Method 7
2.2.1 Notation 7
2.2.2 Landmarks on the Histogram 7
2.2.3 Overview of the Standardizing Method 9
2.2.4 Algorithms 9
2.3 Variants of the Standardizing Method 11
2.3.1 Choosing the Standardization Parameters 12
2.3.2 Choosing the Landmark Configuration 17
2.3.3 Choosing the Mapping Function 18
2.4 Evaluation 18
2.4.1 Qualitative Comparison 19
2.4.2 Quantitative Comparison 25
2.5 Concluding Remarks 30

3 Fuzzy-connected 3D Image Segmentation at Interactive Speeds **33**
3.1 Introduction to Image Segmentation 33
3.2 Fuzzy Connectedness Principles 35
3.2.1 Fuzzy Adjacency and Affinity 35
3.2.2 Fuzzy Connectedness and Fuzzy Connected Objects 36
3.2.3 Relative Fuzzy Connectedness 37
3.3 Algorithms 37
3.3.1 Graph Search Algorithms 38
3.3.2 Affinity Computation 38
3.3.3 Tracking 40
3.4 Experiments, Results, and Discussion 43
3.4.1 Image Data and Task 43
3.4.2 Comparisons 46
3.5 Concluding Remarks 51

4 An MRI Protocol-independent Brain Segmentation Method **53**
4.1 Introduction to Image Segmentation Methods 53
4.1.1 Segmentation Methods in MRI 54
4.1.2 Need for Image Segmentation Workshops 55
4.2 Interactive Segmentation Method 55
4.2.1 Overview of the Segmentation Method 55
4.2.2 Training Phase 56

4.2.3 Correcting for RF Field Inhomogeneity . . . 56
4.2.4 Standardizing MR Image Intensities . . . 57
4.2.5 Computing Fuzzy Affinity . . . 57
4.2.6 Specifying Seed Points and Volume of Interest . . . 57
4.2.7 Computing Fuzzy Connectedness . . . 58
4.2.8 Creating the Brain Mask . . . 58
4.2.9 Verifying the Brain Mask . . . 58
4.2.10 Creating Masks for Each Object and Computing Volume . . . 58
4.3 Evaluation . . . 58
4.3.1 Image Data . . . 58
4.3.2 Precision (Reproducibility) . . . 59
4.3.3 Accuracy . . . 59
4.3.4 Efficiency . . . 60
4.4 Results and Discussion . . . 60
4.4.1 Display Examples . . . 60
4.4.2 Precision (Reproducibility) . . . 61
4.4.3 Accuracy . . . 64
4.4.4 Efficiency . . . 65
4.5 Concluding Remarks . . . 65

5 A Task-specific Comparison of 3D Image Registration Methods 67
5.1 Introduction to Image Registration . . . 67
5.1.1 Registration as Applied to Studying MS . . . 68
5.2 The Compared Registration Methods . . . 69
5.2.1 General Scheme for Registration . . . 70
5.2.2 Scale as a New Feature . . . 70
5.2.3 Transformation . . . 72
5.2.4 Matching Criteria . . . 73
5.2.5 Initialization . . . 74
5.2.6 Search Strategy . . . 75
5.3 Evaluation . . . 75
5.3.1 Image Data . . . 75
5.3.2 Experiments . . . 76
5.3.3 Figures of Merit . . . 77
5.4 Results and Discussion . . . 78
5.5 Concluding Remarks . . . 80

6 MR Image Analysis in Multiple Sclerosis 83
6.1 Introduction to the Study of Multiple Sclerosis . . . 83
6.2 Frameworks for MR Image Analysis in MS . . . 84
6.3 T2-weighted Imaging . . . 85
6.3.1 Thresholding Techniques . . . 86
6.3.2 Parametric/Non-parametric Classification, Clustering . . . 87
6.3.3 Miscellaneous Advanced Methods . . . 89
6.4 T1-weighted Imaging with Contrast Enhancement . . . 90
6.4.1 Methods Used with T1WIE . . . 90
6.5 Magnetization Transfer Imaging . . . 91
6.5.1 MTR Analysis Methods . . . 91
6.6 T1-weighted Imaging . . . 92
6.6.1 Analysis Techniques for T1WI . . . 93
6.7 Fuzzy Connectedness Methods . . . 93

6.7.1 Data Acquisition 95
6.7.2 Analysis of FSE T2, PD Images 95
6.7.3 Analysis of T1WIE Images 96
6.7.4 Analysis of MT Images 96
6.8 Evaluation of the Fuzzy Connectedness Methods 97
6.8.1 Validation 97
6.8.2 Clinical Correlation 97
6.9 Discussion and Concluding Remarks 98
6.9.1 Comparative Assessment 99
6.9.2 Automation, Failure, User Assistance 99
6.9.3 Standardized MR Intensity-based Analysis 100
6.9.4 Segmentation “Workshop” 100

Bibliography **101**

Summary **119**

Összefoglaló **125**

Publications **131**

Chapter 1

Introduction

This thesis deals with issues from different aspects of medical image processing—image segmentation, image registration, and image analysis. Although the techniques and algorithms are described within the context of an actively pursued application, namely the analysis of brain magnetic resonance images (MRI) in Multiple Sclerosis, most of the methods described here are applicable to other brain diseases and other modalities as well. They even have potential in non-medical applications.

Image Processing

Image processing, whether it be applied to robotics, computer vision, or medicine, will treat imaging geometry, linear transforms, the frequency domain, digital and continuous domains, segmentation, histogram analysis, morphology, and other topics that apply to any imaging modality and any application. Image analysis, regardless of its application area, encompasses the incorporation of prior knowledge, the classification of features, the matching of models to images, the description of shape, and many of the generic problems and approaches of artificial intelligence. However, while these classic approaches to general images and to general applications are important, the special nature of medical images and medical applications requires special treatment.

Medical Image Processing

The special nature of medical images derives as much from their method of acquisition as it does from the subjects whose images are being acquired. The development of medical imaging over the past three decades has been revolutionary. With the increasing availability of relatively inexpensive computational resources, computerized techniques such as computed tomography (CT), magnetic resonance imaging (MRI), digital subtraction angiography, Doppler ultrasound imaging, and various imaging techniques based on nuclear emission (PET, SPECT, etc.) have all been valuable additions to the radiologist's arsenal of imaging tools toward ever more reliable detection and diagnosis of disease. Data sets in two, three, or more dimensions convey increasingly vast and detailed information for clinical or research applications. This information has to be interpreted in a timely and accurate manner to benefit health care.

Medical image processing deals with the development of problem-specific approaches to the enhancement of raw medical image data for the purposes of selective visualization as well as further analysis. Medical image analysis then concentrates on the development of techniques to supplement the mostly qualitative and frequently subjective assessment of medical images by human experts with a variety of new information that is quantitative, objective, and reproducible.

What is perhaps most remarkable about these advances in medical imaging is the fact that the challenges have required significant innovation in computational techniques for nearly all aspects of

image processing in various fields. The use of multiple imaging modalities on a single patient, for example MRI and PET, requires sophisticated algorithms for image registration and pattern matching. Automated recognition and diagnosis require image segmentation, quantification and enhancement tools. Algorithms for image segmentation and visualization are employed broadly through many applications using all of the digital imaging modalities. The aim behind all these efforts is that by carefully tailoring and combining these image processing tools we create additional aids to the physicians in diagnosis, treatment planning, and disease monitoring and, thus, ultimately benefit patient care.

Standardization of the MRI Intensity Scale

A major difficulty with the MRI techniques for most protocols has been that image intensities do not have a fixed meaning, even within the same protocol, for the same body region, for images obtained on the same scanner, for the same patient. This poses problems in finding the proper window setting for image display, and more importantly, in image segmentation and analysis. Without the same protocol-specific intensity meaning, setting values for the parameters of image segmentation methods in a patient-study-independent fashion becomes very difficult. This results in compromising precision, accuracy, and efficiency of segmentation.

We recently developed an image processing method for MRI intensity standardization wherein all images (independent of patients and the specific brand of the MR scanner used) can be transformed in such a way that for the same protocol and body region, in the transformed images similar intensities will have similar tissue meaning. The method is described and extensively evaluated in Chapter 2.

Image Segmentation

In brain MRI analysis, usually the first step is to segment the brain parenchyma and remove the extrameningeal tissues and bone. Also, in many situations, separating the gray matter and white matter is desirable, as well as separating normal and abnormal tissues such as normal appearing and abnormal normal appearing white matter, and lesions.

Image segmentation is the process of identifying and delineating objects in images. Segmentation is needed directly or indirectly for most of the operations done on images. It is also the most difficult of all image operations. Segmentation may be thought of as consisting of two related tasks—recognition (or detection) and delineation. In most recognition tasks, trained human operators outperform any computer algorithms. On the contrary, computer algorithms exist for delineation that are more precise, accurate, and efficient than human delineation of object regions or boundaries.

In automatic segmentation methods, a question arises as to what to do in case of failures. The premise in interactive methods is that, often a simple help from an operator is sufficient as a recognition aid. In clinical practice, the methods utilized in clinical trials invariably require some human assistance for each study for a human reader to check the results to insure that they are correct. Therefore, if we can make this process efficient, the uncertainties of the automatic methods can be overcome, and we have a solution that is practical.

The graded composition of tissues resulting from pathological heterogeneity and partial volume effects are not accurately accounted for by binary delineation methods. Fuzzy image processing frameworks become essential to handle material heterogeneity, noise, background intensity variation, fuzziness inherent in images, and how image elements hang together in image space and feature space in the presence of this fuzziness.

Image segmentation techniques using fuzzy connectedness principles have been used successfully in several large applications in recent years: MR brain image analysis for quantifying white matter lesions and brain tumors, segmentation of vessel structures and artery-vein separation in

MRA, craniofacial muscle segmentation, and mammographic fibroglandular density quantification. However, there were several parts in these techniques where improvements can be achieved.

When the fuzzy connectedness algorithms were first introduced they could not be used for interactive processing due to their excessive computational requirements. In an attempt to substantially speed up the fuzzy connectedness methods we studied systematically a host of 'optimal' graph search algorithms. In Chapter 3 we demonstrate, by extensively testing these algorithms on a variety of 3D medical images taken from large on-going applications, that substantial improvement is achievable with a combination of well-chosen algorithms and fast modern PCs.

Despite the vast amount of literature and wide range of approaches to image segmentation, general, reusable frameworks seem to be lacking. Most papers deal with methods devised for specific tasks, tailored to a specific acquisition protocol and body region. What we need is a segmentation "workshop" wherein a protocol-specific segmentation method can be quickly fabricated. That is, for a new protocol, the workshop allows quick experimentation to complete the set up phase quickly rather than requiring months and years of basic research and development.

We devised one such workshop for brain MRI segmentation. In Chapter 4, we present a method that combines the robust, accurate, and efficient techniques of fuzzy connectedness segmentation with standardized MRI intensities and fast algorithms. The result is a general segmentation framework that efficiently utilizes the user input (for recognition) and the power of computer (for delineation). This same two-phase method has been applied to segment brain tissues from a variety of MRI protocols.

Image Registration

To be able to utilize the complementary information from multiprotocol acquisitions, or for longitudinal image analysis, the images need to be in spatial correspondence. Image registration is used to achieve this correspondence. Intra-subject image registration is used for matching multiple acquisitions for the same study or for matching longitudinal acquisitions. Inter-subject registration is used when creating atlases (of "average" anatomy), or when using the atlases in individual cases (e.g., for matching the atlas to the study in question and classifying the tissues based on the pre-determined classification of the atlas).

The various MR imaging methods provide somewhat disparate pieces of the information relating to the MS puzzle. The use of quantitative MRI in predicting the onset of MS and in assessing disease severity in established MS is actively pursued at present. In all such efforts, image registration to combine the complementary information from the images acquired under different protocols becomes vital. Many of the image segmentation methods for MS require one or more preprocessing operations, including registration to handle multichannel features. New quantitative measures are also being developed from the images for each of several tissue regions including the whole brain parenchyma, gray matter, white matter, lesions and cerebrospinal fluid (CSF), for patients with MS. These methods also depend on image registration.

In Chapter 5, we present a new class of approaches for rigid-body registration. We compare their performance with several commonly used strategies in studying Multiple Sclerosis via multi protocol MRI.

Image Analysis in Multiple Sclerosis

Multiple Sclerosis (MS) is an acquired disease of the central nervous system (CNS). Since its first description in 1877, it has been investigated extensively. In its advanced state, the disease may severely impair the ambulatory ability and may even cause death. There are several clinical measures commonly used to express the severity of the disease. These measures are subjective and may have reproducibility problems. The development of new treatments demands objective outcome measures for relatively short trials. MR imaging has proven to be a very sensitive marker of the MS disease. A

variety of protocols are being investigated for improving the conspicuity of the lesions in the images in the early stage and in the advanced stage for both the microscopic and macroscopic processes.

A review of MR image analysis techniques currently used in MS is given in Chapter 6. Also, the family of techniques using fuzzy connectedness for different protocols is described and evaluated.

Chapter 2

Standardizing the MR Image Intensity Scale

One of the major drawbacks of Magnetic Resonance Imaging (MRI) has been the lack of a standard and quantifiable interpretation of image intensities. Unlike in other modalities such as x-ray computerized tomography, MR images taken for the same patient on the same scanner at different times may appear different from each other due to a variety of scanner-dependent variations, and therefore, the absolute intensity values do not have a fixed meaning. We devised a two-step method wherein all images (independent of patients and the specific brand of the MR scanner used) can be transformed in such a way that for the same protocol and body region, in the transformed images similar intensities will have similar tissue meaning. In the first step, the parameters of the standardizing transformation are "learnt" from an image set. In the second step, for each MR study, these parameters are used to map their histogram into the standardized histogram. The method was tested quantitatively on 90 whole brain studies of Multiple Sclerosis patients for several protocols and qualitatively for several other protocols and body regions. Measurements using mean squared difference showed that the standardized image intensities have statistically significantly ($p < 0.01$) more consistent range and meaning than the originals. This consistency achieved seems to be independent of the scanners. Standardized images can be displayed with fixed windows without the need of per case adjustment. Our results also show that the method facilitates improving the degree of automation of image segmentation. More importantly, extraction of quantitative information about healthy organs or about abnormalities can be considerably simplified. In this chapter, we introduce the standardizing method and compare its variants.

The standardizing framework is described in Sections 2.2 and its variants are given in 2.3. We conduct quantitative comparisons of intensities for the methods utilizing different configurations of landmarks, as described in Section 2.4. We state our concluding remarks in Section 2.5. The content of this chapter appeared in [122, 125, 126, 127, 128, 130, 131, 135].

2.1 The Problem with MR Image Intensities

Magnetic Resonance Imaging (MRI) has revolutionized radiological imaging of the internal structures of the human body. It has the advantage of being noninvasive with no known health hazards. A variety of MRI protocols are currently available, with and without the use of contrast agents, such as T1, T1 with a contrast agent, T2 and proton density (PD) with spin-echo (SE) or fast spin-echo (FSE) sequences, magnetization transfer (MT), FLAIR, SPGR, and GRASS. These protocols allow the setting up of different contrasts among the different tissues within the same organ system.

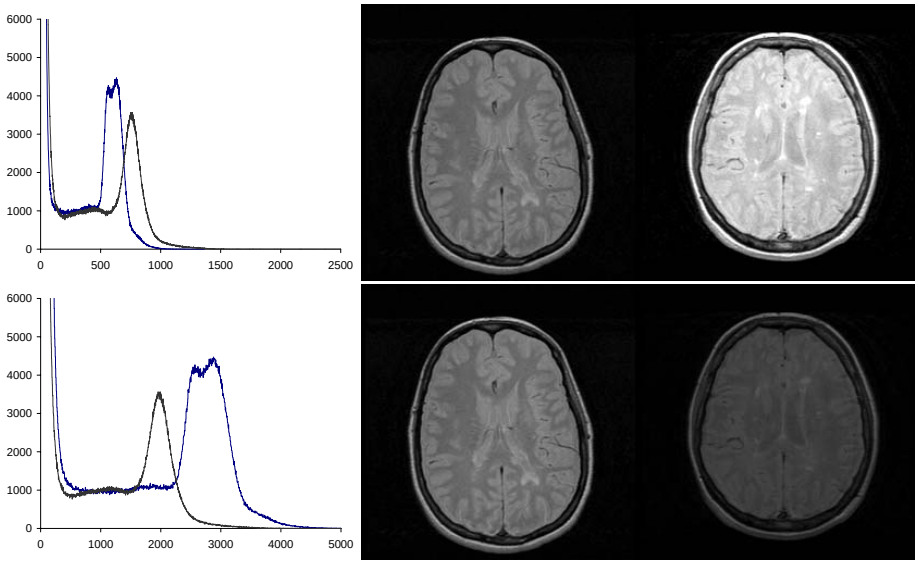


Figure 2.1: A pair of brain images obtained using identical PD protocols on the same scanner, and their histograms (top row), and those images after scaling to a fixed range, and their histograms (bottom row). The window level for display is set at the mid point of the full range and the window width is the full range.

Ironically, this richness of acquisition schemes comes with a major difficulty. The image intensities in MRI do not have a fixed meaning, not even within the same protocol for the same body region obtained on the same scanner for the same patient. This implies that MR images cannot be displayed at preset windows; one may have to adjust the window settings per case. The lack of a meaning for intensities also poses problems in image segmentation [13, 24] and quantification [90, 164, 189].

Most visualization and analysis methods have parameters. The exceptions are perhaps manual methods wherein the human knowledge can be thought of as representing the parameters. However, the results of segmentation by two physicians are likely to differ because of the differences in their training. Setting values for the parameters for the non-manual methods becomes more difficult without the same protocol-specific intensity meaning. What we need is that for protocols that are the same or "close" to each other, the resulting images should be "close".

Attempts have been made in the past to calibrate MR signal characteristics at the time of acquisition using phantoms [38, 214]. Although it is feasible to do such a calibration of all patient scans, it is somewhat cumbersome. Moreover, such a technique is not applicable to image data that have already been acquired without the required calibration phantoms. Postprocessing techniques that are applied to the image data that do not have any special acquisition requirements are clearly more attractive. Such methods would not only make acquiring images simpler but also make it possible to process the already acquired data.

There is a natural tendency to think that a simple scaling of the minimum to maximum intensity range of the given image to a fixed standard range may solve this problem. We illustrate in Fig. 2.1 that this usually does not help in achieving a similarity of intensities.

A postprocessing technique to automatically adjust the contrast and brightness of MR images (i.e., "windowing") for image display has been presented in [207]. However, although such automatic windowing may achieve display uniformity, they may not be adequate for quantitative image analysis, since the intensities still may not have tissue-specific meaning after the windowing transformation. There does not seem to have been any serious attempt to address this problem in the past.

The method described in this chapter offers a simple way of transforming the images so that there is a significant gain in similarity of the resulting images. It is a two-step process consisting of a training step (executed only once for each protocol and body region) and a transformation step (executed on each given image). This new transformation results in a standard scale for each protocol and body region. Intensities in the transformed images have consistent tissue meanings, and standard window settings can be determined for different tissues just as in CT.

The main idea underlying the methods is to deform the image histograms so that they match a mean histogram determined through training. The actual matching is based on certain landmarks identified on the histograms. In the original method [128], we utilized a low and a high percentile point together with the mode corresponding to the foreground in the image as the landmarks. Although, for display, this achieves a remarkably consistent window level and width setting, for image segmentation, the mode-based matching is often too sensitive to the actual location of the mode which is often quite variable. In a later work [135], we described variants of the method that utilized other landmarks, the quartiles and deciles, in particular the median corresponding to the foreground in the image, in addition to the low and high percentiles, to overcome the difficulties with the mode. The method is easy to implement and rapid in execution. The actual transformation itself can be stored as a lookup table in the image header. The transformed images permit predetermined display window settings and also facilitate image segmentation.

2.2 The Standardizing Method

In this section, we will describe our standardizing method. The framework will be set up in a general way so that both the earlier method [128] and the new variants [135] can also be described in the same framework.

2.2.1 Notation

We denote the set of MRI protocols by $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$ and the set of body regions by $\mathcal{D} = \{D_1, D_2, \ldots, D_m\}$. We represent a *volume image* (or an *image* for short) by a pair $\mathcal{V} = (V, g)$, where V is a 3-dimensional array of volume elements (voxels) covering a body region of the particular patient for whom image data $\mathcal{V}$ are acquired, and g is a function, called *intensity function*, that assigns an integer *intensity value* for each $v \in V$. (We will often refer to a volume image simply as an image for short. This is not to be confused with the 2D slice images. All processing operations described in here are carried out in 3D.) We assume that $g(v) \geq 0$ for all $v \in V$, and $g(v) = 0$ if, and only if, there are no measured (and computed) data for voxel v. We denote by $\mathcal{V}_{P,D}$ the set of all images that can possibly be generated as per protocol $P \in \mathcal{P}$ for body region $D \in \mathcal{D}$.

The *histogram* of any image $\mathcal{V}$ is a pair $\mathcal{H} = (G, h)$ where G is the set of all possible intensity values (gray values) in $\mathcal{V}$ (i.e., the range of g), and h is a function whose domain is G and whose value for each $x \in G$ is the number of voxels $v \in V$ for which $g(v) = x$. Let $m_1 = \min\{\, g(v) \mid v \in V \text{ and } g(v) > 0 \,\}$ and $m_2 = \max\{\, g(v) \mid v \in V \text{ and } g(v) > 0 \,\}$, the minimum and maximum intensity values in $\mathcal{V}$, respectively.

2.2.2 Landmarks on the Histogram

It is desirable to cut off the tails of the histogram of the image because they often cause problems. Usually the high intensity tail corresponds to artifacts and outlier intensities, and causes considerable inter- and intra-patient/scanner variations. With this in mind, let pc_1 and pc_2 denote the minimum and maximum percentile values, respectively, that are used to select a range of *intensity of interest* (IOI). Let the actual intensity values corresponding to pc_1 and pc_2 in the histogram $\mathcal{H}$ be p_1 and p_2.

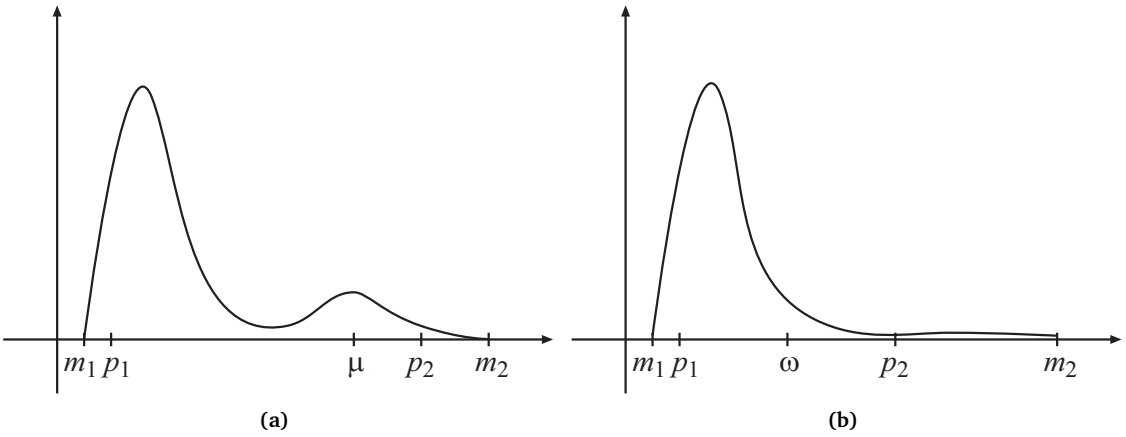


Figure 2.2: Location of the histogram-specific parameters (landmarks). m_1 and m_2 are the minimum and maximum intensities in the image, p_1 and p_2 are the minimum and maximum percentile intensities, μ is the second mode of the histogram in the bimodal case (a) and ω is the shoulder of the "background hump" in the unimodal case (b).

The following relationship holds.

$$100 \cdot \sum_{x=m_1}^{p_i} h(x) \;=\; pc_i \cdot \sum_{x=m_1}^{m_2} h(x) \qquad (i = 1, 2)\,. \tag{2.1}$$

We will come back later to how pc_1 and pc_2 are determined.

Based on examining over 20 body region/protocol combinations, we observed mainly two types of histograms among MR images: unimodal and bimodal (see Fig. 2.2). In case of bimodal histograms, we used in the past the second mode (μ) that corresponds to the main foreground object in the image as a histogram landmark. With unimodal histograms the mode usually corresponds to the background so we need to select some other landmark. This may be for example the shoulder (ω) of the hump of the background intensities (identified by, for example, the point at which the histogram slope becomes -1). The locations of the histogram-specific parameters in these two cases are illustrated in Fig. 2.2 schematically. Since all of the protocols we studied in [128, 135] and in our on-going applications produce bimodal histograms, we will confine our description only to such protocols. With some additional work, the methods can be extended to the unimodal case as well as to other (multimodal) cases.

We will consider the following landmark configurations.

$$\begin{aligned} L_1 &= \{pc_1, pc_2, \mu\}\,, \\ L_2 &= \{pc_1, pc_2, \mu_{50}\}\,, \\ L_3 &= \{pc_1, pc_2, \mu_{25}, \mu_{50}, \mu_{75}\}\,, \\ L_4 &= \{pc_1, pc_2, \mu_{10}, \mu_{20}, \ldots, \mu_{90}\}\,, \end{aligned} \tag{2.2}$$

where μ_p, for $p \in \{10, 20, \ldots, 90, 25, 75\}$, represents the pth percentile in the histogram associated with the foreground part of a given image and μ represents its second mode. For any image $\mathcal{V}_i$, we will denote by p_{1i} and p_{2i} the image intensities corresponding to the percentiles pc_1 and pc_2, respectively, of the histogram of $\mathcal{V}_i$. We will use the notation $\mu_{1i}, \mu_{2i}, \ldots, \mu_{li}$ to denote the image intensities in $\mathcal{V}_i$ corresponding to the landmarks other than pc_1 and pc_2 that are derived from $\mathcal{V}_i$ from among those defined in Eq. (2.2). For example, for the set L_3, $\mu_{1i}, \mu_{2i}, \mu_{3i}$ represent the intensities corresponding to the 25th, 50th and 75th percentiles of the histogram of the foreground of $\mathcal{V}_i$, respectively.

2.2.3 Overview of the Standardizing Method

The method consists of two steps. In the first ("training") step, a set of images of the same body region and protocol corresponding to a population of patients is given as input. The parameters of a histogram transformation are "learnt" from these image data and a few additional input parameters are determined. These additional parameters may vary and are used to fine tune the transformation to the protocol, the body region, and possibly the application that will work on the transformed data. This step needs to be executed only once for a given protocol and for a given body region. In the second ("transformation") step, the images are transformed using the parameters learnt in the first step. This transformation is image dependent and needs to be done for each given image. This step usually results in a nonlinear intensity transformation for the given image (since the segments are mapped independently). However, the relationship between tissue intensities is maintained and intensity comparisons can be made using the standardized images.

Our overall approach is as follows. Let the minimum and the maximum intensities on the standard scale for the IOI be s_1 and s_2, respectively. In the training step, the landmarks (p_{1j}, p_{2j}, μ_{1j}, μ_{2j}, ..., μ_{lj}) obtained (as described later) from each image $\mathcal{V}_j$ of a set of images are mapped to the standard scale by mapping the intensities from $[p_{1j}, p_{2j}]$ onto $[s_1, s_2]$ linearly (see Fig. 2.3). The formula for mapping $x \in [p_{1j}, p_{2j}]$ to x' is the following:

$$x' = s_1 + \frac{x - p_{1j}}{p_{2j} - p_{1j}}(s_2 - s_1)\,. \tag{2.3}$$

Then, for each k, $1 \le k \le l$, the mean μ_{ks} of the mapped μ_{kj}'s over the images $\mathcal{V}_j$ is computed (see Algorithm 1). In the transformation step, for any given image $\mathcal{V}_i$, the actual landmark locations μ_{ki} obtained from its histogram are matched to μ_{ks} by doing several separate linear mappings (see Algorithm 2): the first from $[p_{1i}, \mu_{1i}]$ to $[s_1, \mu_{1s}]$, the second from $[\mu_{1i}, \mu_{2i}]$ to $[\mu_{1s}, \mu_{2s}]$, ..., and the last from $[\mu_{li}, p_{2i}]$ to $[\mu_{ls}, s_2]$, as illustrated in Fig. 2.4. In the past, we used the configuration denoted L_1, with the corresponding transformation that results for a given image illustrated in Fig. 2.4 (a). Figure 2.4 (b) demonstrates the general case of l landmarks in addition to p_{1i} and p_{2i}. In Fig. 2.4, s'_{1i} and s'_{2i} represent the actual extreme values on the standard scale. They are determined by the linear mappings of the extreme segments: $[p_{1i}, \mu_{1i}]$ to $[s_1, \mu_{1s}]$ and $[\mu_{li}, p_{2i}]$ to $[\mu_{ls}, s_2]$. We call the overall mapping from the intensities $[m_{1i}, m_{2i}]$ of $\mathcal{V}_i$ to $[s'_{1i}, s'_{2i}]$ of the standard scale the *standardizer of* $\mathcal{V}_i$, and denote it by $\tau_{\mathcal{V}_i}$. The expression for $\tau_{\mathcal{V}_i}$ (from Fig. 2.3) is as follows:

$$\tau_{\mathcal{V}_i}(x) = \begin{cases} \left\lceil \mu_s + (x - \mu_i)\dfrac{s_1 - \mu_s}{p_{1i} - \mu_i} \right\rceil, & \text{if } m_{1i} \le x \le \mu_i\,, \\ \left\lceil \mu_s + (x - \mu_i)\dfrac{s_2 - \mu_s}{p_{2i} - \mu_i} \right\rceil, & \text{if } \mu_i \le x \le m_{2i}\,, \end{cases} \tag{2.4}$$

where $\lceil \cdot \rceil$ denotes the "ceiling" or "floor" operator. (These operators convert any real number y to the closest integer Y such that $Y \ge y$ or $Y \le y$, respectively.) Note that $s'_{1i} = \tau_{\mathcal{V}_i}(m_{1i})$, and $s'_{2i} = \tau_{\mathcal{V}_i}(m_{2i})$. The image $\mathcal{V}_{si} = (V_i, g_{si})$ resulting from the standardizing mapping of $\mathcal{V}_i$ is given by, for all $v \in \mathcal{V}_i$,

$$g_{si}(v) = \tau_{\mathcal{V}_i}(g_i(v))\,. \tag{2.5}$$

2.2.4 Algorithms

The training and transformation algorithms are presented in Algorithms 1 and 2. Both algorithms are straightforward, and require no special data structures or optimization in implementation.

We point out that the "free ends" characterized by the values of s'_{1i} and s'_{2i} of the standard scale depend on the given image $\mathcal{V}_i$. In other words, the range $[s'_{1i}, s'_{2i}]$ may vary from image to image.

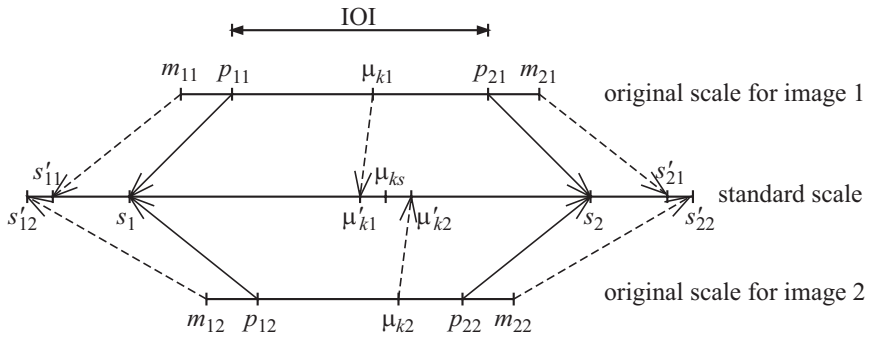


Figure 2.3: Finding the parameters of the standard histogram. For illustration, only two input images are shown. For $j = 1, 2$, m_{1j} and m_{2j} are the minimum and maximum intensities in the image $\mathcal{V}_j$, p_{1j} and p_{2j} are the minimum and maximum percentile intensities, μ_{kj} is one of the landmarks of the histogram, μ'_{kj} is the mapped value of μ_{kj}, and μ_{ks} is the mean of the μ'_{kj}'s—the actual parameter we are looking for on the standard scale.

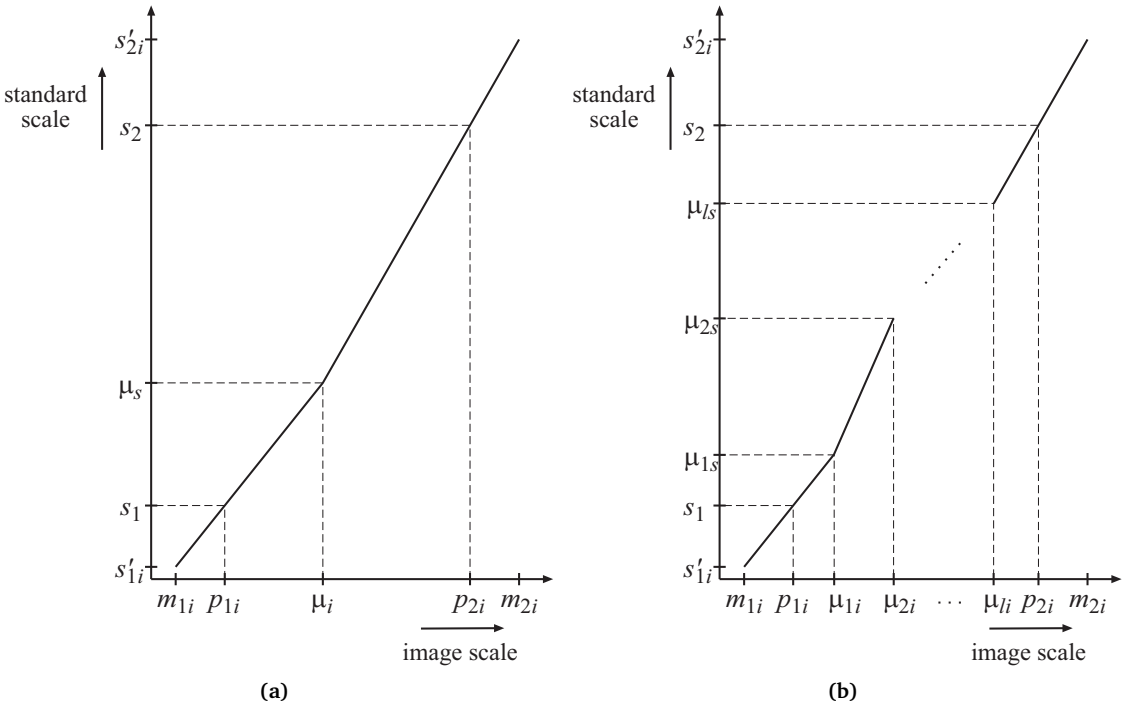


Figure 2.4: The intensity mapping function for the transformation phase. (a) for 3 landmarks (for configuration L_1), (b) for the general case with l landmarks.

Algorithm 1: Training for the standardizer

Input: A set of images $\mathcal{V}_j$ $(j = 1, 2, \ldots, N)$ that is a subset of $\mathcal{V}_{P,D}$ for some fixed $P \in \mathcal{P}$ and $D \in \mathcal{D}$, the histogram parameters pc_1, pc_2, and s_1, s_2, and $L \in \{L_1, L_2, L_3, L_4\}$.

Output: $\{\mu_{ks} \mid 1 \leq k \leq l\}$.

begin

1 **for** $j = 1$ *to* N **do**

2 compute the histogram $\mathcal{H}_j$ of $\mathcal{V}_j$;

3 determine intensity values p_{1j} and p_{2j} corresponding to pc_1 and pc_2 and the landmark locations μ_{1j}, μ_{2j}, ..., μ_{lj} on $\mathcal{H}_j$ (see comments in the text);

4 map $[p_{1j}, p_{2j}]$ of $\mathcal{H}_j$ onto $[s_1, s_2]$ linearly;

5 find the new mapped landmark locations μ'_{1j}, μ'_{2j}, ..., μ'_{lj};

endfor

6 calculate the rounded means μ_{1s}, μ_{2s}, ..., μ_{ls} of μ'_{1j}, μ'_{2j}, ..., μ'_{lj}, respectively, over $j = 1, 2, \ldots, N$;

end

Algorithm 2: Transforming by the standardizer

Input: An image $\mathcal{V}_i \in \mathcal{V}_{P,D}$, pc_1, pc_2, s_1, s_2, μ_{1s}, μ_{2s}, ..., μ_{ls}.

Output: The transformed image $\mathcal{V}_{si}$ and/or a lookup table (LUT) that stores the standardizer $\tau_{\mathcal{V}_i}$.

begin

1 compute the histogram $\mathcal{H}_i = (G_i, h_i)$ of $\mathcal{V}_i$;

2 determine intensity values p_{1i} and p_{2i} corresponding to pc_1 and pc_2 and the landmark locations μ_{1i}, μ_{2i}, ..., μ_{li} on $\mathcal{H}_i$ (see comments below);

3 map sections of the scale of $\mathcal{H}_i$ linearly according to Fig. 2.4 to the standard scale G_s of the standard histogram $\mathcal{H}_s = (G_s, h_s)$;

4 map the intensity value of every voxel $v \in \mathcal{V}_i$ according to $\tau_{\mathcal{V}_i}$ to get $\mathcal{V}_{si}$ and/or output the mapping $\tau_{\mathcal{V}_i}$ in a LUT;

end

However, $[s_1, s_2]$ is independent of $\mathcal{V}_i$, and this is the interval within which a uniformity of intensity meaning is achieved.

In order to find $\mu_{1i}, \mu_{2i}, \ldots, \mu_{li}$ for a given image $\mathcal{V}_i$, we need to identify roughly the foreground part of the image. In an image of a head of a patient, for example, this part corresponds roughly to the set of all voxels that fall in the head. One of several simple methods can be utilized for this purpose. We employ thresholding. The threshold is chosen to be the overall mean intensity in the whole image. The method of background removal described in [189] can also be used. Several hundred studies have been processed automatically and successfully using this method in our ongoing projects for different protocol and body region combinations.

2.3 Variants of the Standardizing Method

In the following sections, we shall give some guidelines for selecting parameters for the standardizer. We shall also describe strategies for constructing different variants of the standardizer.

2.3.1 Choosing the Standardization Parameters

Although once the training step is done the corresponding transformation step is fully determined, there are several possibilities to tailor the standardizer to the specific needs of an application. For example, s_1 should not be 0 if the values below p_1 need to be distinguished from "nothing" (i.e., value 0). Further, s_2 should be large enough not to compress the parts of the histograms (i.e., not to merge neighboring intensities after the transformation). When $pc_1 > 0$ and/or $pc_2 < 100$, the values in $[m_{1i}, p_{1i}]$ and $[p_{2i}, m_{2i}]$ must be mapped to $[s'_{1i}, s_1]$ and $[s_2, s'_{2i}]$, respectively, where s'_{1i} and s'_{2i} are determined by applying the mapping in the two linear sections of Fig. 2.4.

We delineated a theory and established a set of theorems for the configuration denoted L_1 that are crucial to guarantee the correct behavior (i.e., preserving the intensity relations between voxels) of the standardizer $\tau_{\mathcal{V}_i}$ for any given image $\mathcal{V}_i$. The theory readily generalizes to configurations L_2, L_3, and L_4.

We will need the following definitions for stating the theorems. Let

$$\mu'_{\min} = \min_{\mathcal{V}_i \in \mathcal{V}_{P,D}} \{\mu'_i\}\,, \tag{2.6}$$

and

$$\mu'_{\max} = \max_{\mathcal{V}_i \in \mathcal{V}_{P,D}} \{\mu'_i\}\,. \tag{2.7}$$

Note that $\mu'_{\min}$ and $\mu'_{\max}$ are dependent on s_1 and s_2. Let l, L, r, and R be indexes such that

$$\mu_l - p_{1l} = \min_{\mathcal{V}_i \in \mathcal{V}_{P,D}} \{(\mu_i - p_{1i})\}\,, \tag{2.8}$$

$$\mu_L - p_{1L} = \max_{\mathcal{V}_i \in \mathcal{V}_{P,D}} \{(\mu_i - p_{1i})\}\,, \tag{2.9}$$

$$p_{2r} - \mu_r = \min_{\mathcal{V}_i \in \mathcal{V}_{P,D}} \{(p_{2i} - \mu_i)\}\,, \tag{2.10}$$

$$p_{2R} - \mu_R = \max_{\mathcal{V}_i \in \mathcal{V}_{P,D}} \{(p_{2i} - \mu_i)\}\,. \tag{2.11}$$

We will assume throughout that pc_1 and pc_2 are such that, for any $\mathcal{V}_i \in \mathcal{V}_{P,D}$,

$$p_{1i} < \mu_i < p_{2i}\,. \tag{2.12}$$

Let x and y be any real numbers, and let $\lceil \cdot \rceil$ be any "rounding" operator (such as, "ceiling" or "floor"). The first Lemma is quite simple, however, because it will be used in several proofs, we state and prove it separately.

Lemma 1. *For any protocol $P \in \mathcal{P}$, any body region $D \in \mathcal{D}$, any image $\mathcal{V}_i \in \mathcal{V}_{P,D}$, and for any pc_1 and pc_2 satisfying Eq.* (2.12)*, if*

$$\mu'_{\min} - s_1 \geq \mu_L - p_{1L}\,, \tag{2.13}$$

and

$$s_2 - \mu'_{\max} \geq p_{2R} - \mu_R\,, \tag{2.14}$$

then

$$\mu_s - s_1 \geq \mu_i - p_{1i}\,, \tag{2.15}$$

and

$$s_2 - \mu_s \geq p_{2i} - \mu_i\,. \tag{2.16}$$

Proof. By Eqs. (2.6), (2.7), and the fact that μ_s is the mean of μ_i's for all $\mathcal{V}_i \in \mathcal{V}_{P,D}$, it follows that $\mu'_{\min} \leq \mu_s \leq \mu'_{\max}$. By Eqs. (2.9) and (2.11), $\mu_L - p_{1L} \geq \mu_i - p_{1i}$ and $p_{2R} - \mu_R \geq p_{2i} - \mu_i$, for any $\mathcal{V}_i \in \mathcal{V}_{P,D}$. Combining these with Eqs. (2.13), and (2.14), Eqs. (2.15) and (2.16) follow. □

Lemma 2. *For any protocol $P \in \mathcal{P}$, any body region $D \in \mathcal{D}$, any image $\mathcal{V}_i \in \mathcal{V}_{P,D}$, and for any pc_1 and pc_2 satisfying Eq.* (2.12)*, if Eqs.* (2.13) *and* (2.14) *are satisfied, and if x_1, x_2 and μ_i are intensities of $\mathcal{V}_i$ such that $x_1 < \mu_i < x_2$, then*

$$\tau_{\mathcal{V}_i}(x_1) < \mu_s < \tau_{\mathcal{V}_i}(x_2)\,. \tag{2.17}$$

Proof. Assume $x_1 < \mu_i$. Then, by Eq. (2.4), $\tau_{\mathcal{V}_i}(x_1) < \mu_s$ can be written as

$$\left\lceil \mu_s + (x_1 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \right\rceil < \mu_s\,, \tag{2.18}$$

which is equivalent to

$$\mu_s - \left(\mu_s + (x_1 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \right) \geq 1\,, \tag{2.19}$$

which can be further simplified to

$$(\mu_i - x_1)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \geq 1\,. \tag{2.20}$$

By assumption, $\mu_i - x_1 > 0$, and since μ_i and x_1 are integers, $\mu_i - x_1 \geq 1$. By Eq. (2.15), $(s_1 - \mu_s)/(p_{1i} - \mu_i) \geq 1$, hence Eq. (2.20) holds, and so $\tau_{\mathcal{V}_i}(x_1) < \mu_s$.

Now assume $\mu_i < x_2$. By Eq. (2.4), $\mu_s < \tau_{\mathcal{V}_i}(x_2)$ can be written as

$$\mu_s < \left\lceil \mu_s + (x_2 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \right\rceil\,, \tag{2.21}$$

which is equivalent to

$$\left(\mu_s + (x_2 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \right) - \mu_s \geq 1\,, \tag{2.22}$$

which can be further simplified to

$$(x_2 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \geq 1\,. \tag{2.23}$$

By assumption, $x_2 - \mu_i > 0$, and since x_2 and μ_i are integers, $x_2 - \mu_i \geq 1$. By Eq. (2.16), $(s_2 - \mu_s)/(p_{2i} - \mu_i) \geq 1$, hence Eq. (2.23) holds, and so $\mu_s < \tau_{\mathcal{V}_i}(x_2)$. □

Theorem 1 states the conditions under which it is guaranteed that no two distinct intensities in $\mathcal{V}_i$ are merged into a single intensity in the standardized image $\mathcal{V}_{si}$. Thus, if standardizing is done respecting these conditions, then there is no loss of information, and the original image can be obtained by inverting the standardizer $\tau_{\mathcal{V}}$.

Theorem 1. *For any protocol $P \in \mathcal{P}$, any body region $D \in \mathcal{D}$, any image $\mathcal{V}_i \in \mathcal{V}_{P,D}$, and for any pc_1 and pc_2, such that Eq.* (2.12) *is satisfied, if Eqs.* (2.13) *and* (2.14) *are satisfied, the standardizer $\tau_{\mathcal{V}_i}$ of $\mathcal{V}_i$ is a one-to-one mapping.*

Proof. Let x_1 and x_2 be intensities of $\mathcal{V}_i$ such that $m_{1i} \leq x_1 \leq x_2 \leq m_{2i}$. We need to consider the following three cases:

$$x_1 \leq x_2 \leq \mu_i\,, \tag{2.24}$$

$$\mu_i \leq x_1 \leq x_2\,, \tag{2.25}$$

$$x_1 < \mu_i < x_2\,. \tag{2.26}$$

Case of Eq. (2.24): $\tau_{\mathcal{V}_i}(x_1) = \tau_{\mathcal{V}_i}(x_2)$ can be written as

$$\left\lceil \mu_s + (x_1 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \right\rceil = \left\lceil \mu_s + (x_2 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \right\rceil , \tag{2.27}$$

which is equivalent to

$$\left| \left(\mu_s + (x_1 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \right) - \left(\mu_s + (x_2 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \right) \right| < 1 , \tag{2.28}$$

which can be further simplified to

$$|x_1 - x_2|\frac{s_1 - \mu_s}{p_{1i} - \mu_i} < 1 . \tag{2.29}$$

By Eq. (2.15), $(s_1 - \mu_s)/(p_{1i} - \mu_i) \geq 1$. To satisfy Eq. (2.29), $|x_1 - x_2| < 1$ must hold, which is equivalent to $x_1 = x_2$, since x_1 and x_2 are integers.

Case of Eq. (2.25): $\tau_{\mathcal{V}_i}(x_1) = \tau_{\mathcal{V}_i}(x_2)$ can be written as

$$\left\lceil \mu_s + (x_1 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \right\rceil = \left\lceil \mu_s + (x_2 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \right\rceil , \tag{2.30}$$

which is equivalent to

$$\left| \left(\mu_s + (x_1 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \right) - \left(\mu_s + (x_2 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \right) \right| < 1 , \tag{2.31}$$

which can be further simplified to

$$|x_1 - x_2|\frac{s_2 - \mu_s}{p_{2i} - \mu_i} < 1 . \tag{2.32}$$

By Eq. (2.16), $(s_2 - \mu_s)/(p_{2i} - \mu_i) \geq 1$. To satisfy Eq. (2.32), $|x_1 - x_2| < 1$ must hold, which is equivalent to $x_1 = x_2$, since x_1 and x_2 are integers.

Case of Eq. (2.26): By Lemma 2, $\tau_{\mathcal{V}_i}(x_1) < \tau_{\mathcal{V}_i}(x_2)$, and there is nothing to prove. □

The following theorem gives a guidance for selecting the values of s_1 and s_2 that cause no intensity loss.

Theorem 2. *For any protocol $P \in \mathcal{P}$, any body region $D \in \mathcal{D}$, any image $\mathcal{V}_i \in \mathcal{V}_{P,D}$, and for any pc_1 and pc_2, such that Eq.* (2.12) *is satisifed, the standardizer $\tau_{\mathcal{V}_i}$ of $\mathcal{V}_i$ is a one-to-one mapping if*

$$s_2 - s_1 \geq (\mu_L - p_{1L} + p_{2R} - \mu_R) \cdot \max \left\{ \frac{\mu_L - p_{1L}}{\mu_l - p_{1l}}, \frac{p_{2R} - \mu_R}{p_{2r} - \mu_r} \right\} . \tag{2.33}$$

Note that Theorems 1 and 2 state conditions that require observing all images in $\mathcal{V}_{P,D}$. In practice, since this is impossible to do, we estimate the right side of Eq. (2.33) by examining a sufficient number of images and set $s_2 - s_1$ to a number sufficiently greater than this estimated entity. Our implemented software gives a warning message should an image be encountered for which this condition is violated. Even in such cases of violation, the software can be used in such way that, when a violation is detected, $s_2 - s_1$ is automatically updated so that this condition is indeed satisfied.

In testing 100 proton density (PD) 3D images of the brain of different patients, acquired as per a fixed fast-spin-echo protocol on two GE 1.5T scanners, we found that $\mu_L - p_{1L} = 1139$, $p_{2R} - \mu_R = 960$, $\mu_l - p_{1l} = 693$, and $p_{2r} - \mu_r = 502$ for $pc_1 = 0$ and $pc_2 = 99.8$. By Eq. (2.33), this implies that, with $s_1 = 1$, if we set s_2 to at least 4016, we can make sure that the mapping is lossless (though probably in many cases smaller value would be sufficent). In all our experiments (and current use), we set $s_2 = 4095$. This choice has been adequate in all protocols we currently

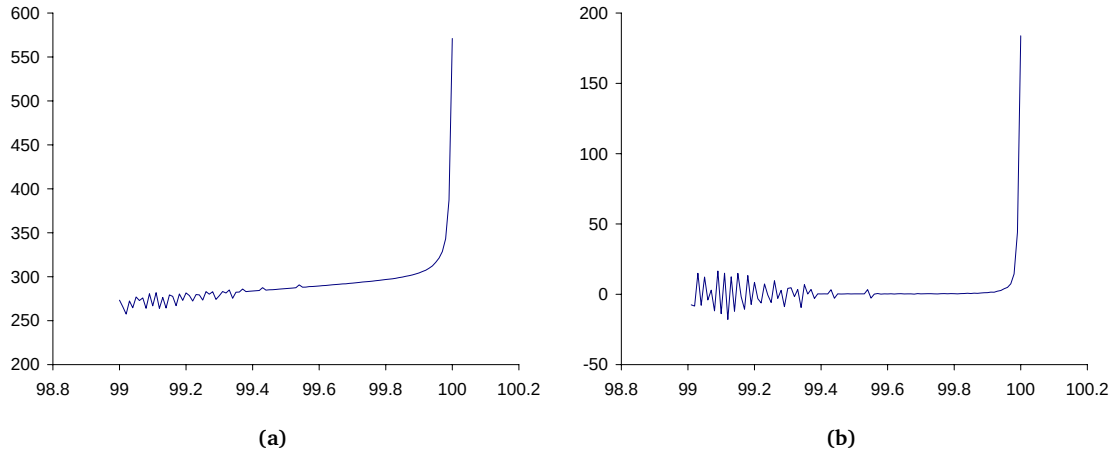


Figure 2.5: Plot of the standard deviation of p_{2i} for the images in the test set versus pc_2 (a), and its first derivative (b). $pc_2 = 99.8$ is approximately the largest value at which the derivative is still small (0.5).

use and is highly unlikely to lead to a lossy mapping at least for any proton density brain images acquired as per the above protocol.

We arrived at the choice of $pc_2 = 99.8$ by the following experiments. We randomly selected a few PD brain images acquired as per the above protocol. With $pc_1 = 0$ fixed, we computed p_{2i} for several values of pc_2 (in steps of 0.05, between 99.0 and 100.0) for each test image $\mathcal{V}_i$. We chose the value of pc_2 to be the largest value at which the variation in p_{2i} with respect to changes in pc_2 reached an acceptably small value. Figure 2.5 shows the plot of the standard deviation of p_{2i} for the images in the test set versus pc_2 Fig. 2.5 (a) and the derivative of that function Fig. 2.5 (b). The value $pc_2 = 99.8$ corresponds to the shoulder in the graph on the right where the function starts increasing rapidly. This implies that beyond this pc_2 value, the variations in the intensities are not systematic but random.

Proof of Theorem 2. Equation (2.33) holds if, and only if,

$$s_2 - s_1 \geq (\mu_L - p_{1L} + p_{2R} - \mu_R)\frac{\mu_L - p_{1L}}{\mu_l - p_{1l}}\,, \tag{2.34}$$

and

$$s_2 - s_1 \geq (\mu_L - p_{1L} + p_{2R} - \mu_R)\frac{p_{2R} - \mu_R}{p_{2r} - \mu_r}\,. \tag{2.35}$$

Using Eqs. (2.9), (2.11), and (2.8), the right side of Eq. (2.34) can be rewritten as

$$s_2 - s_1 \geq (\mu_i - p_{1i} + p_{2i} - \mu_i)\frac{\mu_L - p_{1L}}{\mu_i - p_{1i}}\,. \tag{2.36}$$

After simplifying and rearranging Eq. (2.36), we arrive at

$$(s_2 - s_1)\frac{\mu_i - p_{1i}}{p_{2i} - p_{1i}} \geq \mu_L - p_{1L}\,. \tag{2.37}$$

Using the definition of x' in Eq. (2.3), and using μ_i in place of x,

$$(s_2 - s_1)\frac{\mu_i - p_{1i}}{p_{2i} - p_{1i}} = \mu_i' - s_1\,, \tag{2.38}$$

and Eq. (2.37) can be written as

$$\mu_i' - s_1 \geq \mu_L - p_{1L} \,. \tag{2.39}$$

Analogously, using Eqs. (2.9), (2.11), and (2.10), the right side of Eq. (2.35) can be rewritten as

$$s_2 - s_1 \geq (\mu_i - p_{1i} + p_{2i} - \mu_i)\frac{\mu_R - p_{1R}}{p_{2i} - \mu_i} \,. \tag{2.40}$$

After simplifying and rearranging Eq. (2.40), we arrive at

$$(s_2 - s_1)\frac{p_{2i} - \mu_i}{p_{2i} - p_{1i}} \geq \mu_R - p_{1R} \,. \tag{2.41}$$

Using the definition of x' in Eq. (2.3), and using μ_i in place of x,

$$(s_2 - s_1)\frac{p_{2i} - \mu_i}{p_{2i} - p_{1i}} = s_2 - \mu_i' \,, \tag{2.42}$$

and Eq. (2.41) can be written as

$$s_2 - \mu_i' \geq \mu_R - p_{1R} \,. \tag{2.43}$$

Since Eqs. (2.39) and (2.43) hold for any $\mathcal{V}_i \in \mathcal{V}_{P,D}$, all conditions in Theorem 1 are met. Therefore, $\tau_{\mathcal{V}_i}$ is a one-to-one mapping for any $\mathcal{V}_i \in \mathcal{V}_{P,D}$. □

Theorem 3 states that, once the conditions of Theorem 1 are satisfied, the order of the input intensities is maintained in the output images.

Theorem 3. *For any protocol $P \in \mathcal{P}$, any body region $D \in \mathcal{D}$, any image $\mathcal{V}_i \in \mathcal{V}_{P,D}$, any standardizer $\tau_{\mathcal{V}_i}$ of $\mathcal{V}_i$ that satisfies Eq.* (2.33)*, and for any intensities x_1 and x_2 of $\mathcal{V}_i$, $\tau_{\mathcal{V}_i}(x_1) < \tau_{\mathcal{V}_i}(x_2)$ if, and only if, $x_1 < x_2$.*

That is, the actual order of brightness of tissue regions in $\mathcal{V}_i$ is maintained in the image $\mathcal{V}_{si}$ output by Algorithm 1, although their relative contrast may change. That is, the intensity characteristics of $\mathcal{V}_{si}$ are in complete agreement with the characteristics of $\mathcal{V}_i$ that we know as being associated with protocol P. (Note that, if $x_1 = x_2$, then $\tau_{\mathcal{V}_i}(x_1) = \tau_{\mathcal{V}_i}(x_2)$ is guaranteed by Eq. (2.4).) In a PD image of a brain, for example, the known brightness relationship gray matter > white matter > CSF is maintained in $\mathcal{V}_{si}$. The relative contrast between these regions, however, may change.

Proof of Theorem 3. Let x_1 and x_2 be any intensities of $\mathcal{V}_i$ such that $m_{1i} \leq x_1 < x_2 \leq m_{2i}$. Since $x_1 < x_2$, and both are integers,

$$x_2 - x_1 \geq 1 \,. \tag{2.44}$$

We need to consider the following three cases:

$$x_1 < x_2 \leq \mu_i \,, \tag{2.45}$$

$$\mu_i \leq x_1 < x_2 \,, \tag{2.46}$$

$$x_1 < \mu_i < x_2 \,. \tag{2.47}$$

Case of Eq. (2.45): We claim that the following inequality holds. (Note that the two expressions on the left side of "≥" are indeed $\tau_{\mathcal{V}_i}(x_1)$ and $\tau_{\mathcal{V}_i}(x_2)$ from (2.4).)

$$\left(\mu_s + (x_2 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i}\right) - \left(\mu_s + (x_1 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i}\right) \geq 1 \,, \tag{2.48}$$

which, after simplification, is equivalent to

$$(x_2 - x_1)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \geq 1 \,. \tag{2.49}$$

By Eq. (2.15), $(s_1-\mu_s)/(p_{1i}-\mu_i) \geq 1$, hence, by Eq. (2.44), the product on the left side of Eq. (2.49) is indeed ≥ 1. Therefore, Eq. (2.49), and the equivalent Eq. (2.48) hold. Eq. (2.48) is equivalent to

$$\left\lceil \mu_s + (x_1 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \right\rceil < \left\lceil \mu_s + (x_2 - \mu_i)\frac{s_1 - \mu_s}{p_{1i} - \mu_i} \right\rceil . \tag{2.50}$$

By Eq. (2.4), Eq. (2.50) is equivalent to

$$\tau_{\mathcal{V}_i}(x_1) < \tau_{\mathcal{V}_i}(x_2) . \tag{2.51}$$

Case of Eq. (2.46): We claim that the following inequality holds. (Note that the two expressions on the left side of "$\geq$" are indeed $\tau_{\mathcal{V}_i}(x_1)$ and $\tau_{\mathcal{V}_i}(x_2)$ from (2.4).)

$$\left(\mu_s + (x_2 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i}\right) - \left(\mu_s + (x_1 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i}\right) \geq 1 , \tag{2.52}$$

which, after simplification, is equivalent to

$$(x_2 - x_1)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \geq 1 . \tag{2.53}$$

By Eq. (2.16), $(s_2-\mu_s)/(p_{2i}-\mu_i) \geq 1$, hence, by Eq. (2.44), the product on the left side of Eq. (2.53) is indeed ≥ 1. Therefore, Eq. (2.53), and the equivalent Eq. (2.52) hold. Eq. (2.52) is equivalent to

$$\left\lceil \mu_s + (x_1 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \right\rceil < \left\lceil \mu_s + (x_2 - \mu_i)\frac{s_2 - \mu_s}{p_{2i} - \mu_i} \right\rceil . \tag{2.54}$$

By Eq. (2.4), Eq. (2.54) is equivalent to

$$\tau_{\mathcal{V}_i}(x_1) < \tau_{\mathcal{V}_i}(x_2) . \tag{2.55}$$

Case of Eq. (2.47): By Lemma 2, $\tau_{\mathcal{V}_i}(x_1) < \tau_{\mathcal{V}_i}(x_2)$. Combine thiw with Eqs. (2.51) and (2.55), the proof is complete.

Conversely, assume $\tau_{\mathcal{V}_i}(x_1) < \tau_{\mathcal{V}_i}(x_2)$. If $x_1 = x_2$, then, by Eq. (2.4), $\tau_{\mathcal{V}_i}(x_1) = \tau_{\mathcal{V}_i}(x_2)$, therefore $x_1 \neq x_2$. If $x_1 > x_2$, then by the "if" part of the theorem (already proven), $\tau_{\mathcal{V}_i}(x_1) > \tau_{\mathcal{V}_i}(x_2)$, therefore $x_1 \not> x_2$. Hence $x_1 < x_2$. □

2.3.2 Choosing the Landmark Configuration

The choice of the actual landmark configuration is also an important factor. The mode-based method (configuration L_1) works fine for several MR protocols and several body regions for several applications but there are cases (and applications) wherein this simple method is not appropriate. For illustration, we will consider the shape of the gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) distributions in FSE PD brain images. Their relative locations vary among studies and even among studies of the same patient. Figure 2.6 shows some histogram shapes all of which were found in the histograms of Multiple Sclerosis (MS) patient brain studies. We recall here that in FSE PD images, GM regions are brighter than WM regions. Mainly the WM and GM distributions combine to give the histogram hump illustrated in Fig. 2.6 while the CSF distribution falls in the lower tail of this hump. The weakness of the mode-based method is that sometimes the mode (the intensity corresponding to the peak, denoted μ_i in Fig. 2.4 (a)) corresponds to GM intensity (Figs. 2.6 (a) and (b)), in other cases it corresponds to WM intensity (Figs. 2.6 (d) and (e)), or it may also correspond to intensities that lie between real GM and WM intensities (Fig. 2.6 (c)). Therefore, when we match the mode to a fixed location on the standard scale, we may match GM in some cases and WM in the others. Because of this "switching" behavior, the mode-based method

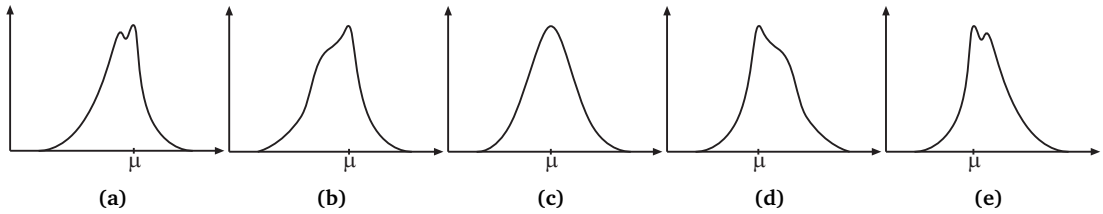


Figure 2.6: Schematic representation of the shapes of brain MRI histograms. For clarity, only the histogram corresponding to the foreground (brain) is shown.

is often not appropriate if the application is image segmentation, where we need more accurate meaning on the standard scale even for relatively small ranges, although for display window setting this behavior does not cause much problems. In order to eliminate the "switching" behavior, one approach is to choose the median of the histogram of the foreground as a landmark to match. By studying about hundred histograms and their landmarks for images of several protocol/body region combinations, we observed that this landmark remains consistent even in cases where the histogram has two similar peaks (Figs. 2.6 (a) and (e)) and an asymmetric shape (Figs. 2.6 (b) and (d)).

We may also use more histogram landmarks, such as quartiles and deciles, to better define the standard histogram. In these percentile-based methods, in addition to the "cut-off" percentiles p_{1i}, p_{2i}, we select several percentile markers $\mu_{1i}, \mu_{2i}, \ldots, \mu_{li}$ within the IOI range. In case of quartiles, $l = 3$ and $\mu_{1i}, \mu_{2i}, \mu_{3i}$ correspond to the 25th, 50th, and 75th percentiles within the IOI range, respectively. In case of deciles, $l = 9$ and $\mu_{1i}, \mu_{2i}, \ldots, \mu_{9i}$ correspond to the 10th, 20th, ..., 90th percentiles, respectively.

2.3.3 Choosing the Mapping Function

There are other possible changes to the basic standardizing method that can be used to make it better fit the type of input data and the actual application: (i) using histogram landmarks defined in different ways (i.e., not by percentiles), (ii) using polynomial functions to "stretch" the histogram segments, (iii) use of spline fitting techniques instead of segment-by-segment linear "stretching". When these variants are used, changes are required in Step 3 of Algorithm 2, wherein the segment-wise linear mapping should be replaced by the actual mapping used.

2.4 Evaluation

Our hypothesis is that, for any given protocol $P \in \mathcal{P}$, and for any body region $D \in \mathcal{D}$, the standardized images $\mathcal{V}_{si}$ have more consistent tissue meaning for image intensities than the images $\mathcal{V}_i$ before standardization. For testing this hypothesis, for each protocol P and body region D, we need to consider the following variations in image data:

(i) intra-patient (time-to-time) variation,

(ii) inter-patient variation,

(iii) variations among different machines of the same brand, and

(iv) variations among machines of different brands.

An evaluation that takes into account all these factors is indeed very formidable. Instead, we will rigorously test the hypothesis for several protocols and for only one body region, namely the brain, and for the factors indicated in only (i) and (ii) above. We will also provide qualitative evidences for the validity of the hypothesis through display examples before and after standardization with different variants for factors indicated in (i), (ii) and (iii) above for several protocols and body

Table 2.1: Parameter configurations used in the comparisons.

method	pc_1	pc_2	s_1	s_2	landmarks
0	—	—	—	—	none
1	0	99.8	1	4095	mode
2	0	99.8	1	4095	median
3	0	99.8	1	4095	quartiles
4	0	99.8	1	4095	deciles

regions. The following sections describe the methods of evaluation that we used to examine how different kinds of variations are handled by the different variants of the standardizing method.

Table 2.1 shows the different parameter settings that were used for the methods in carrying out the comparisons. We refer to the methods by numbers. Zero identifies the identity transformation (i.e., no transformation at all). First, display examples are illustrated with images before and after the different transformations for several protocols. Then quantitative studies are described with illustrating tables. For all tests, segment-wise linear mapping was used. The training was done by using 10 different patient studies of the particular protocol and body region.

2.4.1 Qualitative Comparison

We conducted qualitative comparisons for the following MRI protocols and body regions: fast spin-echo (FSE) PD, FSE T2, spin-echo (SE) PD, SE T2, T1 with Gadolinium enhancement (T1E), and an SPGR sequence, all for the brain; a T1-weighted gradient-echo sequence for the foot. Thirty studies each of FSE PD, FSE T2 and T1E, and 10 studies each of SE PD, SE T2, T1 GRE, and SPGR were transformed using the corresponding "trained" parameters. Within any of these protocols, the image acquisition parameters were identical for all patient studies. Images of all protocols had bimodal histograms except SPGR, which had unimodal histograms. In the latter case, a different landmark set was used in the standardizer (see Fig. 2.2 (b)). All studies used in obtaining the results reported here consisted of 45–60 contiguous 3 mm thick axial slices of size 256×256 with 0.86 mm×0.86 mm in-plane resolution. The voxel intensities were represented as 12 bit integers. No additional preprocessing was performed on any of these images. We also experimented with studies of different slice thickness and orientation and found no significant differences in the results. Since the method is applied to the whole volume histograms and not to individual slices, the slice orientation and the resolution has negligible effect on the transformation. Three types of visual comparisons were made: by plotting histograms of sets of images of the same protocols and body regions, by displaying at fixed gray level window settings, and by displaying the binary images obtained at fixed threshold ranges.

Histograms

Histograms of 10 FSE PD studies at different stages of the transformation are displayed in Fig. 2.7. The low intensity part of the histograms that corresponds to the background voxels has been removed from the display in order to show the intensity of interest on a better scale. Original histograms of the images are plotted in Fig. 2.7 (a). Figure 2.7 (b) displays the histograms of the image after a simple linear mapping of $[p_{1i}, p_{2i}]$ to $[s_1, s_2]$, with $pc_1 = 0$ and $pc_2 = 99.8$. Histograms after the final standardizing mapping are plotted in Fig. 2.7 (c). A visual comparison shows that the histograms are more similar in shape and locations after standardization than before. This implies that

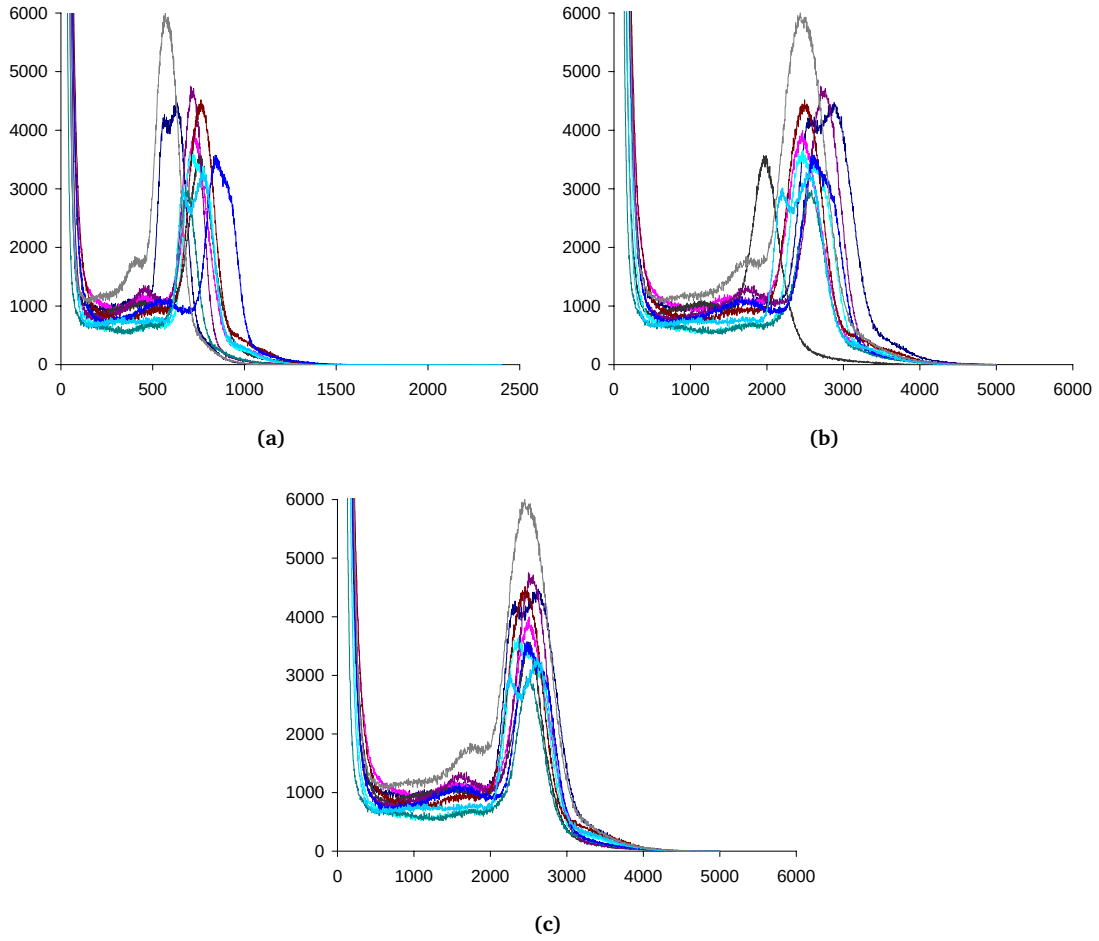


Figure 2.7: Histograms at different stages of the standardization process for 10 different FSE PD studies. Original histograms (a), histograms after intensity scaling from $[p_{1i}, p_{2i}]$ to $[s_1, s_2]$, with $s_1 = 1$, $s_2 = 4095$ (b), and after final standardization (c).

the actual intensity values and their distribution in the images are more similar after standardization. Note also that the simple linear mapping illustrated in the graph in the middle that accounts for outliner intensities is also not enough in making the histograms cluster.

Display at Fixed Windows

Images in each row of Fig. 2.8 show a slice from each of three different patient PD studies. The rows from top to bottom correspond to methods 0 through 4. For each method, the images are displayed at a fixed window setting. For method 0, the window was set up correctly (visually) for the first image. For other methods, we devised "standard" window settings, as listed in Table 2.2, by examining a few standardized studies. This figure demonstrates, that the structures are well portrayed and the contrast is more similar for methods 1–4 than for method 0. Since the intensity ranges are quite different for the three studies, the displays for method 0 exhibit inadequate window settings for the second and third study. Note also the striking similarity of the displays among methods 1–4.

Figures 2.9 and 2.10 are analogous to the first two rows of Fig. 2.8. They show images before and

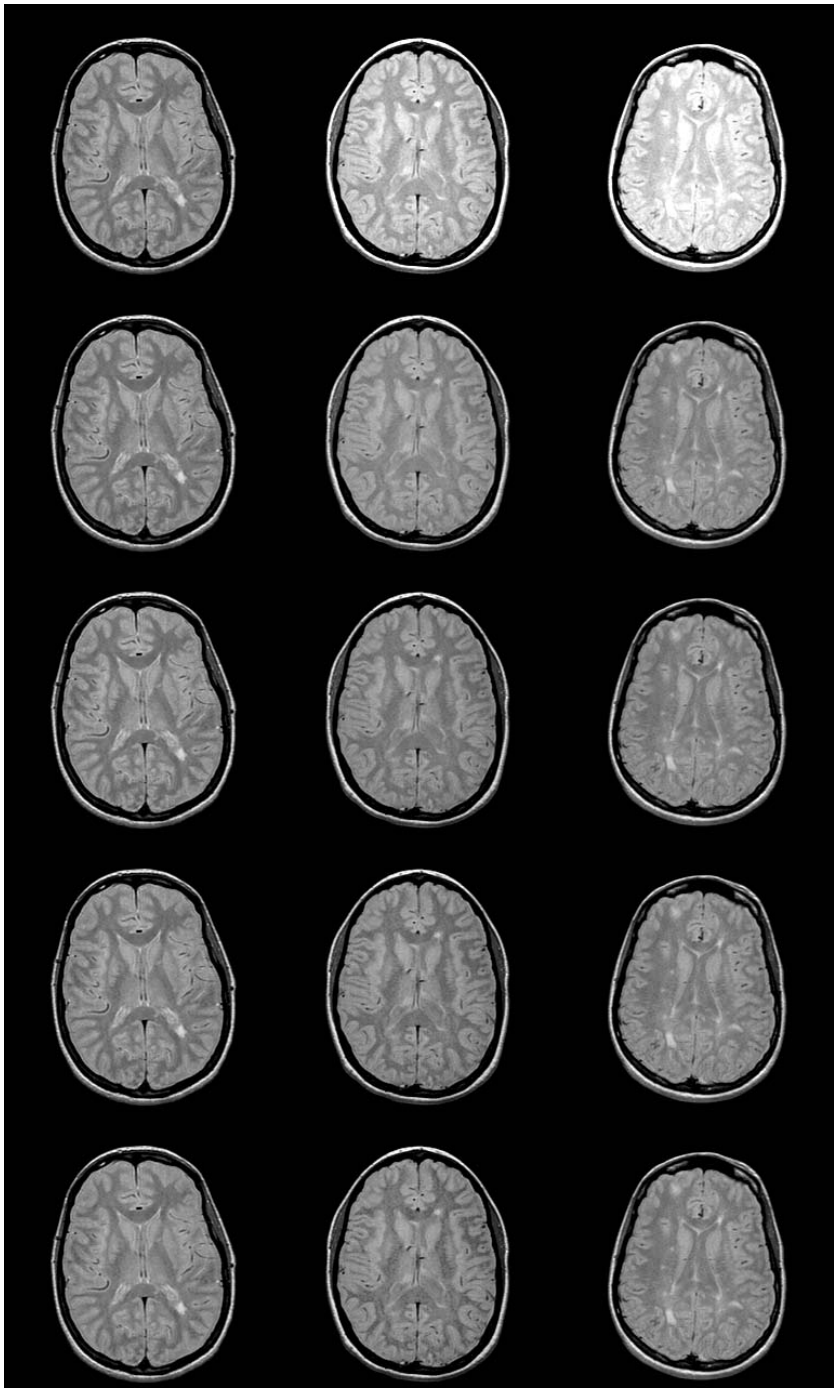


Figure 2.8: Images displayed at fixed gray level windows. Rows 1–5 correspond to methods 0–4. In each row, a slice from the PD studies of three MS patients is displayed.

Table 2.2: Standard windows for the brain for different protocols and methods. Similar tables can be constructed for other protocols and body regions.

method	FSE PD		FSE T2		T1E	
	level	width	level	width	level	width
1	2360	3444	2306	4523	2048	3500
2,3,4	2460	3444	2446	4523	2048	3500

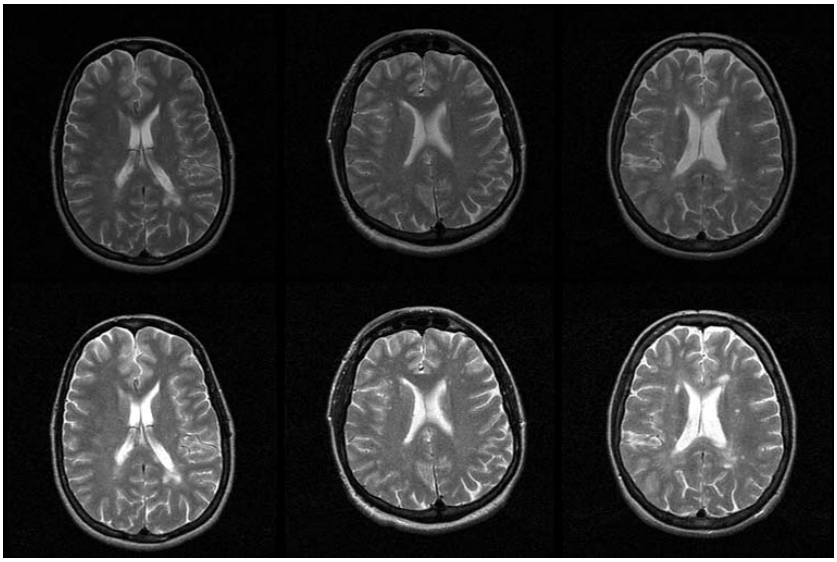


Figure 2.9: Original slices from three studies acquired as per identical FSE T2 protocols before standardization displayed at default windows (first row), and after standardization displayed at a fixed "standard" window (second row).

after standardization for T2 and T1E images, each set acquired using the same protocol. The first rows show the original data displayed using the default window setting (i.e., window level set to the middle of the full range, and window width set to the full range of intensities of the study). The second rows show the same slices after standardization by method 1 displayed using the "standard" brain window settings from Table 2.2.

Figure 2.11 illustrates the improvements resulting from standardization for a different body region, namely the foot. The imaging protocol in this case was a T1-weighted gradient-echo sequence and was the same for the three studies.

The following example is included to demonstrate that the standardizer still works when substantial deviations exist in the test images compared to the images used for training. Figure 2.12 shows a data set of the brain of a patient with a large tumor. The first row of Fig. 2.12 shows a slice of a PD and a T2 data set before standardization and the second row shows the same slices after standardization. In this case, the standardizer was arrived at from the training data utilized for the MS patient PD and T2 images in Figs. 2.8 and 2.9, respectively.

In Fig. 2.13, we illustrate that the method is tolerant to small variations in protocol settings as well as variations that may exist among machines of the same brand in different hospitals. The top row shows three different patient studies obtained from a GE 1.5T Signa scanner at the University of

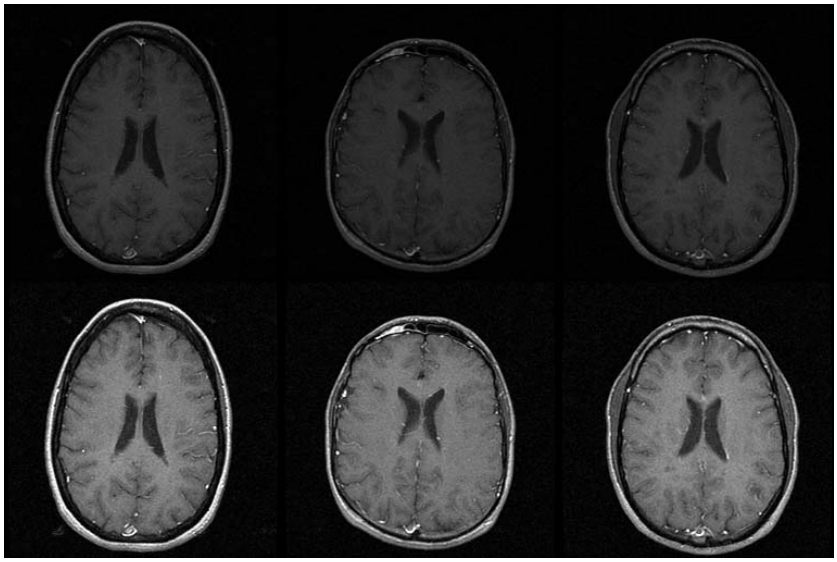


Figure 2.10: Original slices from three studies acquired as per identical T1E protocols before standardization displayed at default windows (first row), and after standardization displayed at a fixed "standard" window (second row).

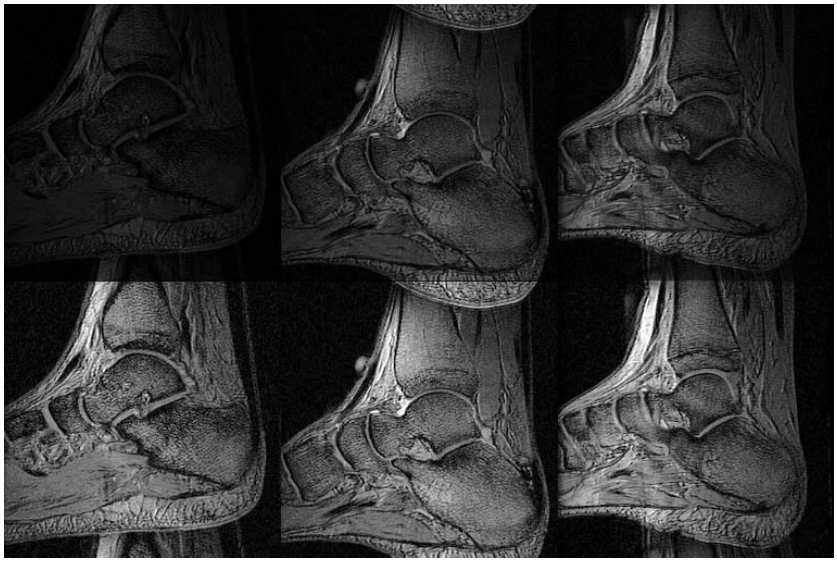


Figure 2.11: Original slices of three foot studies before (first row), and after standardization (second row). The imaging protocol for all three data sets was a T1-weighted gradient echo sequence with identical parameters.

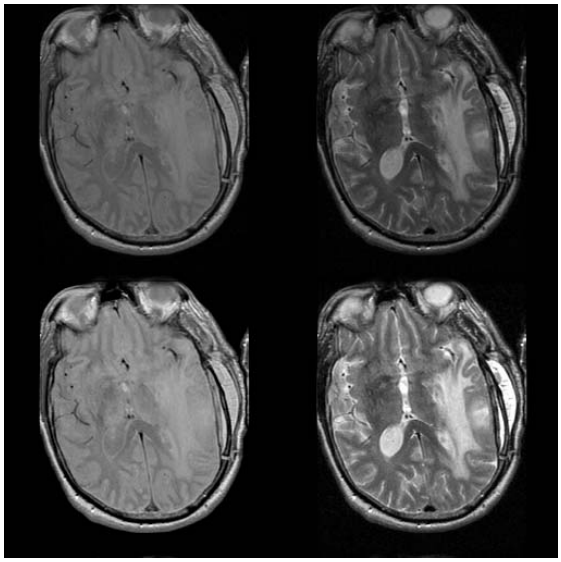


Figure 2.12: Original slices, before standardization, of FSE PD (left) and T2 (right) data sets of a patient's brain with a large tumor displayed at default windows (first row), and after standardization, displayed at a standard window setting (second row) listed in Table 2.2 for the FSE PD and T2 case.

Colorado Health Sciences Center (data courtesy of Dr. Jack Simon) using a SE PD protocol. Images before standardization (displayed using the default window settings) are shown in the first row. The second row shows the same slices after standardization displayed by using the "standard" window setting from Table 2.2. In this case, training was done utilizing the SE PD image data acquired on a scanner of the same brand in our hospital (University of Pennsylvania Health System).

Cross and Mixed Training

The example illustrated in Fig. 2.14 shows that the method is robust against small variations in protocol settings. The first row of Fig. 2.14 shows three SE PD studies before standardization (displayed with the default window setting). The second row shows the same slices after standardization with the parameters that were derived from a training set of an identical SE PD protocol. These images are displayed with the "standard" window setting for SE PD listed in Table 2.2. The third row shows the same slices after standardization with the parameters that were derived from a training set of images acquired as per a FSE PD protocol. These images are displayed with the "standard" window setting for FSE PD listed in Table 2.2. The main difference between the two mappings is a shift of the second mode. This is silently corrected for display when the proper window setting is applied. After both transformations and by using the corresponding window settings, the images look similar and have good brightness and contrast.

The training based on a set of images of a given protocol P and body region D makes the standardizer tuned tightly to the images in $\mathcal{V}_{P,D}$. This is due to the fact that the variation among histograms of images in $\mathcal{V}_{P,D}$ is less than that of images taken from $\mathcal{V}_{P,D}$ and $\mathcal{V}_{P'D'}$ where $P \neq P'$ and/or $D \neq D'$. However, images of different protocols and body regions can be mixed for training and still create a standarizer that fosters consistency of tissue meaning. This is illustrated in Fig. 2.15 wherein images from four different combinations of protocol and body region were utilized in the training step and in devising the standardizer. The figure quite elegantly demonstrates that the standardized images show a better consistency of brightness and contrast at fixed window displays

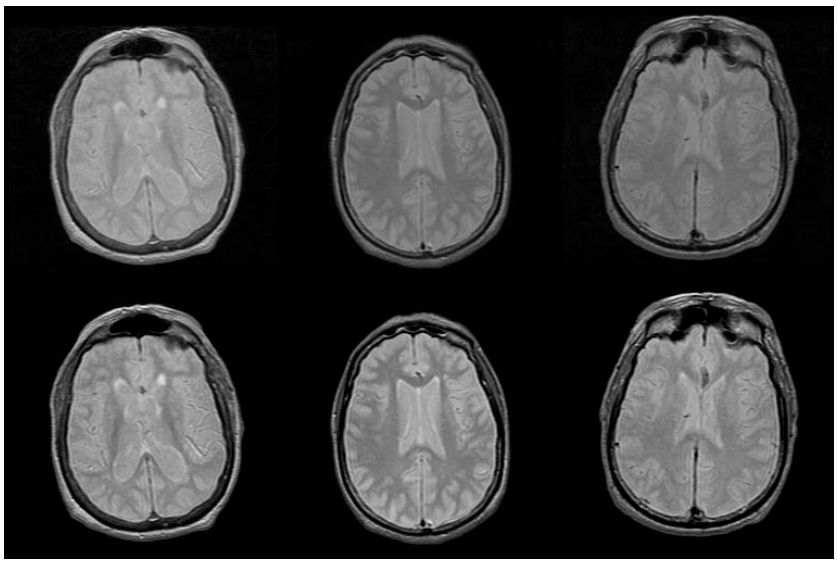


Figure 2.13: Three SE PD studies acquired as per identical protocols before standardization at default windows (first row), and after standardization displayed at a standard window (second row). The training data were acquired as per a similar protocol (with slightly different parameters) from a scanner of the same brand at a different hospital.

(bottom row) than the original images displayed at a default window for each image (top row). This observation begs the question whether it is possible to device a whole body standardizer which will obviate the need for protocol- and body-specific standardizers. We will not pursue this investigation here.

Display as Binary Images at Fixed Thresholds

Figure 2.16 demonstrates in another way that the intensity values have no consistent meaning in the original images but they do have after the transformation. Images are displayed as binary using fixed thresholds to segment approximately the GM region of the brain. These images correspond exactly to those in Fig. 2.8. For each row, the threshold intervals were chosen to roughly segment the GM region in the first study by visual inspection. In row 1, it is well demonstrated that the same thresholds do not segment the same tissue in different studies. In the second study, the given interval corresponds to mainly WM, and in the third, it is far below the brain tissue intensities. The "switching" problem of the mode method is well illustrated by the images in the second row. Although the transformed images displayed at a "standard window" show similar brightness and contrast characteristics, the fixed thresholds for small tissue intensity ranges are not quite applicable. The histograms of the three images in this case are like those shown in Figs. 2.6 (d), (b), and (e). The fixed interval thresholds segment GM in all cases, for methods 2–4 more consistently than in rows 1 and 2. This implies that the meaning of intensities on the standard scale is more consistent after median-based and other percentile-based transformations.

2.4.2 Quantitative Comparison

In order to assess the effectiveness of the standardization method objectively, we conducted two types of quantitative tests on data sets of brain obtained from three protocols: FSE PD, FSE T2, and

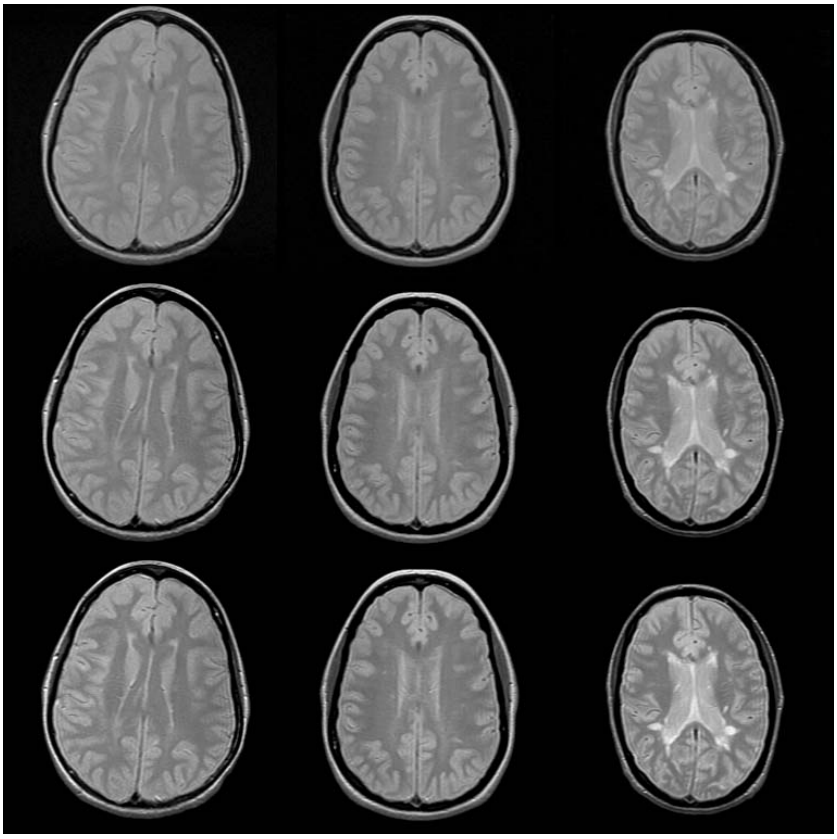


Figure 2.14: Original slices from three studies acquired as per identical SE PD protocols before standardization displayed at default windows (first row), the same slices displayed at a fixed "standard" window after standardization with the parameters that were derived from a training set of an identical SE PD protocol (second row), and after standardization with the parameters that were derived from a training set of images acquired as per a FSE PD protocol displayed at the "standard" window for FSE PD (third row).

T1E. The first test is for assessing intra-patient variation before and after standardization. In the second test, we used an algorithm [189] to segment different tissue regions and compared intensity statistics in different segmented tissue regions for assessing inter-patient variation before and after standardization.

Intra-patient Variation

We used the same training data sets and parameter configurations as for the qualitative comparison described in the previous section. The test method for all three protocols was the same. Two scans acquired at different time instances were randomly selected for 15 patients from our database of MS patient studies. The time distance between the two scans of the same patient varied between 1 and 6 years. For each patient, we registered the first scan to the second via rigid transformation based on intensity value correlations [140]. (See Chapter 5 for an application-specific comparison of image registration methods.) All original images had the same spatial resolution and the paired images covered the same volume of interest. The correlation-based registration method we used proved to result in registration of upto a voxel accuracy. Because these patients had MS and the images of the same patient were acquired 1 to 6 years apart, the number and volume of lesions changed as the

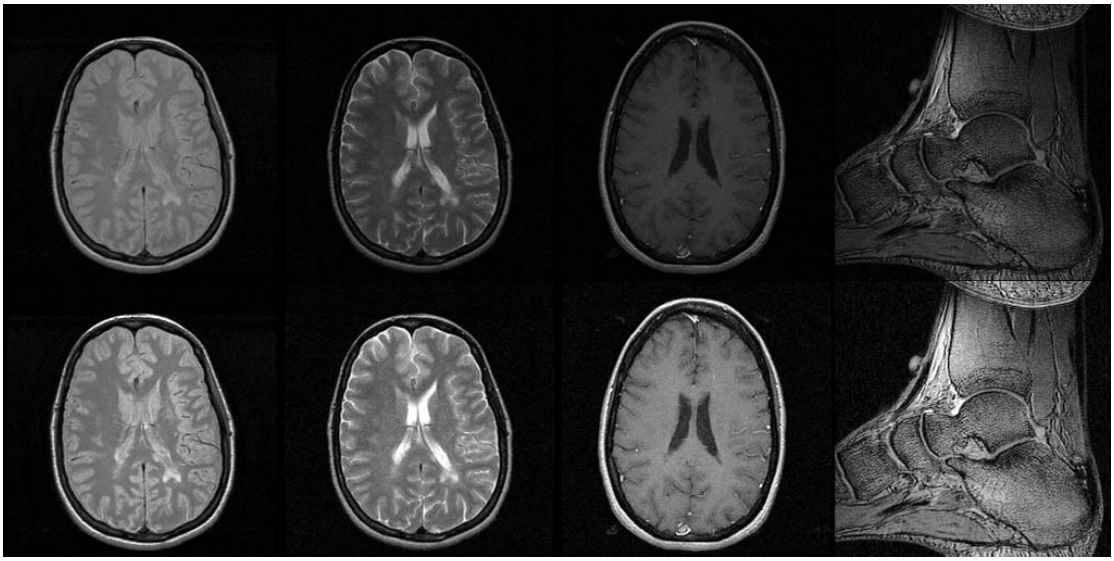


Figure 2.15: Illustration of the effect of mixed training. 10 brain studies from each of FSE PD (first column), FSE T2 (second column), and T1E (third column), and 10 T1-weighted gradient-echo studies of the foot (fourth column) were used in training. The first row shows the display at a study-specific default window of one slice of a study from each of these protocols before standardization. The second row shows the same slices after standardization displayed at one fixed window.

disease evolved. The variation of voxel intensities within these lesion regions is due to the change of the tissue characteristics caused by the disease rather than to the acquisition process. To eliminate these variations in the intensities and to compare only those voxels that supposedly contain the same tissue (mixture), in the two images of the same patient, the lesions were segmented [189] and removed for the purpose of the comparison. However, for the training process, no segmentation was done and the whole volume histograms were used. The lesions in these MS patient studies occupy only a small fraction of the volume of the study and the lesion voxels have high intensity values that are mostly above the 99.8 percentile value, i.e., outside the IOI range used by the method. This small volume of high intensity lesion voxels does not significantly change the median and other landmarks.

The similarity of a pair of these registered, lesion-removed images was measured by the mean squared intensity difference, normalized to the original range of the images, and denoted NMSD. This similarity measure was computed for every pair of images for each method. Table 2.3 shows that the mean value and the standard deviation of the NMSD after each transformation are smaller than that before transformation. The mean values of NMSD for the pairs of studies were compared using the paired t-test under the null hypothesis that there is no difference in NMSD before and after standardization. The hypothesis was rejected at a significance level of $p < 0.01$ indicating that the change in NMSD is statistically significant. The results displayed in Table 2.4 show that the change in the means of NMSD due to the standardizer is statistically significant for all variants (first block). However, the differences in the improvements among the mode-based and other percentile-based variants are not statistically significant (second block). We encountered several patient studies (with histograms as illustrated in Figs. 2.6 (a) and (e)) where the mode-based standardizer failed yet the median-based standardizer succeeded. These "exceptional" cases are as important in practice as the non-exceptionals. If the method breaks down on one of these cases, effort has to be made to specially treat the study in order to deal with the problem yet to retain the overall procedure applied to the

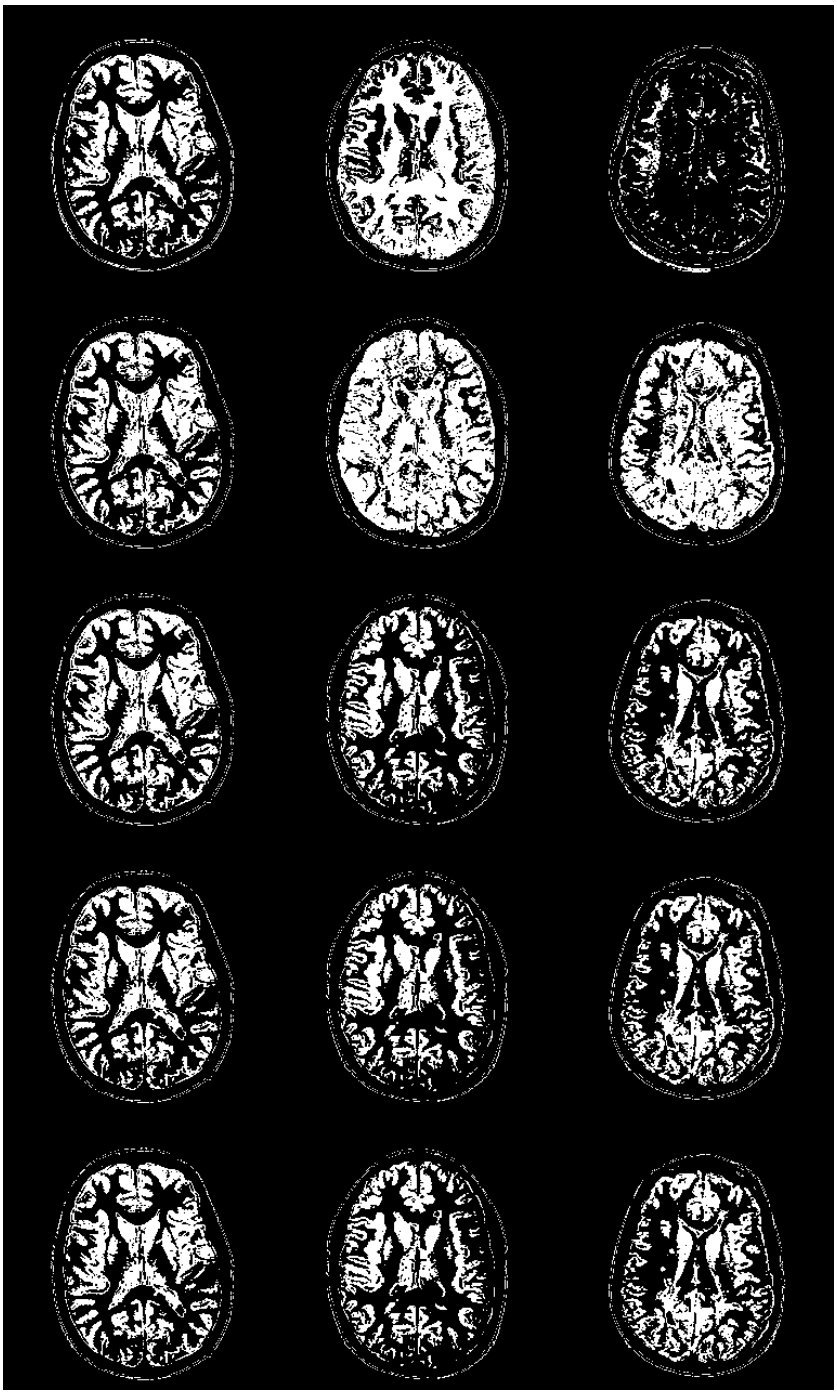


Figure 2.16: Images in Fig. 2.8 displayed as binary using a fixed threshold interval for each row. The threshold interval for the first image in each row was determined interactively to give a good GM definition by visual inspection. Rows 1–5 correspond to methods 0–4.

Table 2.3: Mean and standard deviation of normalized mean squared differences (NMSD) before (method 0) and after different standardizing transforms (methods 1–4) for 15 pairs of studies and for three different protocols. Each pair represents the studies obtained for the same patient at two time instances.

method	FSE PD		FSE T2		T1E	
	mean	sd	mean	sd	mean	sd
0	0.0099	0.0094	0.0093	0.0085	0.0025	0.0018
1	0.0039	0.0055	0.0036	0.0050	0.0020	0.0018
2	0.0039	0.0057	0.0035	0.0050	0.0020	0.0018
3	0.0040	0.0057	0.0033	0.0045	0.0020	0.0018
4	0.0039	0.0054	0.0028	0.0036	0.0019	0.0016

Table 2.4: p-values of paired t-tests applied to the NMSD data.

method	method	FSE PD	FSE T2	T1E
0	1	0.0094	0.0036	0.0450
0	2	0.0101	0.0036	0.0374
0	3	0.0108	0.0031	0.0373
0	4	0.0102	0.0027	0.0241
1	2	0.1722	0.0040	0.5772
1	3	0.0154	0.0396	0.8138
1	4	0.7285	0.0795	0.1862
2	3	0.3415	0.0744	0.5628
2	4	0.6217	0.1007	0.2068
3	4	0.3450	0.1202	0.1925

"normal" cases as much as possible (e.g., being able to apply the same postprocessing steps with the same parameter settings to all studies). The median-based standardizer seems to handle those cases that were "exceptional" for the mode-based method yielding more robust methods. The practical significance of this should not be overlooked yet this is difficult to capture within the statistical mold where the "exceptional one" usually gets lost among the "many normals". Had we included a sufficient number of such exceptional cases in our analysis, we would have seen a statistically significant difference between methods 1 and the rest.

Inter-patient Variation

For this comparison we randomly selected 12 FSE PD and 12 FSE T2 data sets from our database. All images were previously segmented into WM, GM, and CSF regions [189]. The lesions were also segmented [189] and excluded from the tissue regions for comparison. (The reason for not including T1E in this evaluation is that the segmentation of WM, GM, and CSF in this protocol is significantly more difficult.) We calculated the intensity statistics over the population of images for each of these tissue regions separately. The standardizing parameters were the same as those for the other comparisons. For each of these regions in each image $\mathcal{V}_i$ in each of these protocols, we calculated the normalized mean intensity (NMI) by dividing the mean intensity in the region by $p_{2i} - p_{1i}$. This was repeated for each set of the transformed images wherein normalization was

done by dividing the mean intensity in the region by $s_2 - s_1$. The standard deviations of the mean values of the NMI values before and after different standardizing transforms are shown in Table 2.5 together with the corresponding 95% confidence intervals and the coefficients of variation. The table indicates that the intensities on the standard scale have more consistent tissue meaning than those on the original scale and that the median-based standardizer outperforms the mode-based method in achieving similar tissue meaning of intensities. Although for CSF (that included the ventricles as well as the peripheral CSF, where the partial volume effect is significant and results in low intensities) the % CV is slightly higher for the median-based method than that for the mode-based, the overall improvement (though probably not statistically significant) is seen in cases of WM and GM. More importantly the "exceptional" cases (as mentioned previously) are handled correctly by the median and percentile based methods.

2.5 Concluding Remarks

We described a method for standardizing MR image intensity scales. Our goal was to devise a method that post-hoc makes an intensity transformation of images that are routinely acquired, without requiring specialized acquisition protocols and calibration phantoms. The standardizing method aims at achieving consistency of intensity meaning of tissues by devising a transformation that is specific to a given MRI protocol and body region. The basic idea is to deform the histogram of a given 3D image so that it matches a "standard" histogram for that protocol-body-region group. The parameters of the standard histogram are learnt in a training step. We provided theoretical guidelines, and a practical demonstration of how to utilize them, for selecting the values of the parameters of the method and proved that lossless intensity transformation and order is guaranteed, if choices are made as per guidelines.

We also described some of the problems with the earlier MRI scale standardization method reported in [128] and introduced new variants of the method that can help to overcome these problems. Median and other percentile values are more robust than the mode, therefore, using these as landmarks results in a more robust standardizer, hence more consistent meaning of intensities and better defined ranges for different tissues on the standard scale. It is possible to set better intensity of interest ranges while still being able to distinguish relevant information at the ends of the scale.

We presented two types of studies to assess the degree of uniformity of intensity meaning achieved after standardization and the effectiveness of the new variants. In qualitative studies, we showed through image display examples that the consistency of the brightness level and contrast of images is considerably improved after standardization. This permits standardizing and fixing "windows" by protocol, body region, and tissue regions. Using the new standardized images in display, the standard windows for the different tissues (not only for the main object itself) can be either automatically applied or manually selected (e.g., from a short list of available windows), hence saving human interaction time on the per-case manual window adjustments that are currently required in visualizing MR images on physician viewing stations. This may also help in filming MR studies with a uniform appearance of the images.

We also showed that the tissues have much more consistent intensity ranges if the new variants of the standardizing method are used. In quantitative studies, we demonstrated that the scanner dependent intra- and inter-patient intensity variations are substantially reduced after standardizing using the new variants. Although the difference between the reduction yielded by the original standardizing method and by the new variants is not statistically significant, the new variants are more robust and this is important in practical applications. In a subjective examination of the image displays from over 100 studies, we did not come across any case where the method seemed to have failed.

The method is simple, easy to implement, rapid in execution, and completely automatic. It can be incorporated in a Picture Archiving and Communication System (PACS) as a DICOM Value of

Table 2.5: The standard deviation, the 95% confidence interval (CI), and % coefficient of variation (CV) of the normalized mean intensity (NMI) values of different tissues in FSE PD and T2 images.

tissue	method	FSE PD			
		sd	CI low	CI high	% CV
WM	0	0.1047	0.6572	0.7757	14.61
	1	0.0113	0.5523	0.5651	2.03
	2	0.0074	0.5728	0.5812	1.29
	3	0.0111	0.5693	0.5819	1.93
	4	0.0106	0.5695	0.5816	1.85
GM	0	0.1144	0.7189	0.8483	14.59
	1	0.0067	0.6230	0.6306	1.07
	2	0.0065	0.6397	0.6470	1.00
	3	0.0045	0.6449	0.6501	0.70
	4	0.0052	0.6458	0.6518	0.81
CSF	0	0.0884	0.5323	0.6323	15.18
	1	0.0172	0.5309	0.5504	3.18
	2	0.0185	0.5470	0.5679	3.33
	3	0.0172	0.5481	0.5675	3.08
	4	0.0185	0.5450	0.5660	3.33
tissue	method	FSE T2			
		sd	CI low	CI high	% CV
WM	0	0.0580	0.3580	0.4236	14.83
	1	0.0091	0.3456	0.3559	2.59
	2	0.0090	0.3455	0.3557	2.56
	3	0.0086	0.3407	0.3504	2.49
	4	0.0099	0.3377	0.3489	2.88
GM	0	0.0643	0.4124	0.4852	14.33
	1	0.0119	0.4502	0.4636	2.60
	2	0.0097	0.4440	0.4549	1.85
	3	0.0137	0.4299	0.4453	3.13
	4	0.0137	0.4268	0.4423	3.15
CSF	0	0.0791	0.4564	0.5458	15.78
	1	0.0348	0.6596	0.6989	5.12
	2	0.0318	0.6572	0.6932	5.15
	3	0.0296	0.6422	0.6757	4.49
	4	0.0259	0.6141	0.6435	4.12

Interest lookup table, so that, images are automatically transformed or accompanied by the correct lookup table when they are downloaded to the viewing station. It can even be built into the MR scanner to automatically produce images with the standard scale.

Our studies (see Chapters 4 and 6) show that image analysis and tissue segmentation methods—our main motivation to undertake this research—are considerably improved in terms of their constancy of parameter settings and their degree of automation upon scale standardization, especially using the median-based method. With standardization, numerical meaning is achieved, and hence numerical diagnosis and study of diseases may become possible. For example, for MS studies, we may be able to specify an interval $[t_1, t_2]$ of the standardized PD values which correspond to normal WM tissue, another interval $[t_3, t_4]$ corresponding to WM that is normal appearing but that is actually affected by the disease. Our preliminary results indicate that when the values of the imaged parameter (T2, PD, etc.) do not intrinsically overlap for the different tissues, then such a numerical characterization may be possible. We use the standardization method as the first step in all MR image visualization and analysis tasks in which we are currently engaged. We recommend that the method be used as the first step in every MR image visualization and analysis task.

Figures 2.14 and 2.15 indicate a possible simplification of our approach. Instead of devising the standardization transform by protocol and body region, it may be possible to devise them based only on the type of the histogram (Fig. 2.2). That is, the protocol- and body-region-specific training and transformation will then be obviated, hopefully achieving consistency of intensity meaning independent of the protocol and body region. If we match the shoulder of the single mode in Fig. 2.2 (b) with the second mode Fig. 2.2 (a) for images with different types of histograms, it may even be possible to devise standardization schemes that are independent of the histogram type. Can this lead to an MR intensity scale that is similar to the CT scale in the spirit of tissues-specific intensity meaning?

Chapter 3

Fuzzy-connected 3D Image Segmentation at Interactive Speeds

Image segmentation techniques using fuzzy connectedness principles have shown their effectiveness in segmenting a variety of objects in several large applications in recent years. However, one problem with these algorithms has been their excessive computational requirements. In an attempt to substantially speed them up, in this chapter, we study systematically a host of 18 'optimal' graph search algorithms. Extensive testing of these algorithms on a variety of 3D medical images taken from large on-going applications demonstrates that a 700–35,500 fold improvement over current speeds is achievable with a combination of algorithms and fast modern PCs. Existing implementation executes 12–36 times faster on a PC with a 1.4 GHz AthlonXP CPU running Linux 2.4 compared to a 50 MHz Sun SPARCstation-20 running Solaris 2.5. Utilizing efficient algorithms and careful selection of implementations can speed up the computation of fuzzy connectedness values by a factor of 16 to 29 (on the same hardware), as compared to the implementation previously used in our applications utilizing fuzzy object segmentation. The optimality of an algorithm depends on the input data as well as on the choice of the fuzzy affinity relation. The running time is reduced considerably (by a factor up to 34 for brain MR and even more for bone CT), when the algorithms make use of pre-determined thresholds for the fuzzy objects. The reliable recognition (assisted by human operators) and the accurate, efficient, and sophisticated delineation (automatically performed by the computer) can be effectively incorporated into a single interactive process. If images having intensities with tissue-specific meaning (such as CT or standardized MR images) are utilized (see Chapter 2), most of the parameters for the segmentation method can be fixed once for all, all intermediate data (feature and fuzzy affinity values for the whole scene) can be computed before the user interaction is needed and the user can be provided with more information at the time of interaction.

The principles of fuzzy connectedness are described briefly in Section 3.2. In Section 3.3, we present the algorithms and the data structures studied and used in achieving the improved speed. The materials utilized, the testing environment, and the results are detailed in Section 3.4. Finally, we state our concluding remarks in Section 3.5. This chapter is based on [123, 124].

3.1 Introduction to Image Segmentation

The purpose of image segmentation is to extract object information from given images and to output this as a structure system. Segmentation is needed directly or indirectly for most of the operations done on images. It is also the most difficult of all image operations.

Segmentation may be thought of as consisting of two related tasks—*recognition* (or detection) and *delineation*. Recognition is the high-level task of determining roughly the whereabouts of the

object in the given image. Delineation is the low-level task of determining the precise spatial extent of the object and its point-by-point graded composition. In most recognition tasks, trained human operators outperform any computer algorithms. On the contrary, computer algorithms exist for delineation that are more precise, accurate, and efficient than human delineation of object regions or boundaries. Human delineation that specifies graded objectness is impossible at present. Recognition and delineation are not completely disparate steps in segmentation.

Approaches to recognition may be broadly classified into two groups: *automatic* and *human-assisted*. In automatic recognition methods (either knowledge-based [66, 170] or atlas-based [62, 121]), a question arises as to what to do in case of failures. The premise in human assisted methods is that, often a simple help from an operator (on a per-study basis) is sufficient as a recognition aid. Therefore, if we can make this process efficient, the uncertainties of the automatic methods can be overcome, and we have a solution that is practical. This help may be in the form of (i) specification of a few "seed" points in the object region or on its boundary, (ii) indication of a box enclosing the object, or (iii) clicking of a mouse button to accept a real object that has been delineated or to reject a false object.

Image segmentation research for delineation spans nearly four decades [139]. Usually delineation itself is considered to be the total segmentation problem. That is, whatever is output by the delineation method is considered to represent the object of interest.

Images are by nature fuzzy. Approaches to object information extraction from images should therefore attempt to retain uncertainties as realistically as possible. Local fuzziness has been accounted for in fuzzy delineation approaches in the past. A recently developed theoretical and algorithmic framework called fuzzy connectedness attempts to capture the notion of "hanging togetherness" of image elements in a global fuzzy sense for object definition [187].

Image segmentation techniques using fuzzy connectedness principles [187] have been used successfully in several large applications (involving over 2000 3D studies) in recent years: MR brain image analysis for quantifying white matter lesions [113, 164, 189], brain tumor quantification [118], segmentation of vessel structures and artery-vein separation in magnetic resonance angiography (MRA) [94, 155], craniofacial muscle segmentation [188], and mammographic fibroglandular density quantification [162]. However, one problem with the fuzzy connected object segmentation algorithms has been their excessive computational requirements. In the brain lesion segmentation application, for example, the segmentation method [189] took approximately one hour per study (on a Sun SPARCstation-20), and most of this time was spent on segmenting fuzzy connected objects. In order to make this robust and effective technique practically more useful, we investigated its bottleneck operations and found several ways for substantial improvements, to the point that the entire segmentation process can be carried out at close to interactive speeds on modern PCs.

We emphasize that the purpose of this chapter is not reporting a new framework for segmentation but to describe ways of significantly improving the efficiency of the already established fuzzy connectedness framework. Since the algorithmic behavior for the more efficient algorithm is the same as that of the slower versions, we are not undertaking, in this chapter, any application-specific evaluation of the precision and accuracy of the algorithms. These have already been reported previously. Our focus in this chapter therefore will be algorithmic speed.

A preliminary effort in this direction was made recently [17] using Dijkstra's algorithm in place of the original dynamic programming algorithm suggested in [187]. Based on 2D phantoms and two 2D MR slice images, this work [17] demonstrated that a 7–8 fold speed up in fuzzy connectedness object extraction can be achieved. Since the segmentation tasks considered in this paper were somewhat artificial and not in 3D, it was not clear how the observed speed up will hold for segmenting real 3D objects in routine practice. Thus, motivated by this result, in the this chapter, we study systematically a host of 18 optimal graph search algorithms on a variety of 3D medical images taken from several large on-going applications.

The performance of the hardware is rapidly improving. Inexpensive, high-performance PC-based image processing "workstations" can achieve a speed up of 36 or more over a slow workstation using

the same software and algorithms. During the 3 years since the time of the original investigation it has changed so much that, in order to include up-to-date figures, we recomputed all results on a current machine. Even some of our statements in the original report [124] had to be revised. The hardware speed up factor was "only" about 6 (using a 300 MHz Pentium II) at the time of the original investigation in 1999. However, in most cases, the hardware improvement alone is still not enough to make the application run at interactive speeds. By selecting better algorithms and carefully setting their parameters, the computation of fuzzy objects can be speeded up substantially even on the same hardware. This is the component on which we shall focus in this chapter.

We demonstrate that up to 29 fold speed up is achievable by wisely selecting the tracking algorithm for the application in an actual routine practice. If we combine the efficient algorithms with the modern fast PC hardware, we can achieve a processing speed fast enough for interactive segmentation applications using fuzzy connectedness. This, we believe, is a major improvement in segmentation. The reliable recognition (assisted by human operators) and the accurate, efficient, and sophisticated delineation (automatically performed by the computer) can be effectively incorporated into a single interactive process. When these algorithms were first introduced they could not be used for interactive processing. Today, by a combination of the algorithm improvements proposed herein and increased hardware efficiency, this class of algorithms can successfully be incorporated in an interactive segmentation process. One such segmentation process is described in Chapter 4.

The optimality of an algorithm depends on the input data as well as on the type of affinity relations used. It is useful to offer several algorithms and versions in the actual implementation and let the user select the most appropriate for the actual data and task at hand. In our experience, for blob-like objects with a spread-out distribution of their intensities or other properties (such as brain tissues in MRI), algorithms with affinity-based hashing are to be selected. For more sparse objects (vessels, bone) with high contrast (MRA, CT), algorithms with geometric hashing are to be chosen. These guidelines are supported by figures in later sections of this chapter.

3.2 Fuzzy Connectedness Principles

Before delving into the details of our algorithms, we explain the main idea behind fuzzy connectedness. To prevent duplication in subsequent chapters, we describe, in this section, all concepts related to fuzzy connectedness that will facilitate easy understanding of this chapter and the next. The concepts described here are applicable to n-dimensional (fuzzy) digital spaces. For the purposes of illustration, we shall consider three-dimensional examples; see [187] for details.

We refer to an n-dimensional digital image as a *scene* and represent it by a pair $\mathcal{C} = (C, \mathbf{f})$, where C is a rectangular n-dimensional array of spatial elements (*spels* for short) and $\mathbf{f}$ is a function (called *scene intensity*) which assigns to every spel a vector of m integer intensity values. Spels are pixels for $n = 2$ and voxels for $n = 3$. We use the notation $c = (c_1, c_2, \ldots, c_n)$ to denote the coordinates of any spel $c \in C$ in the array. Throughout this thesis, in the context of practical 3D examples, we shall use voxel and spel interchangeably.

3.2.1 Fuzzy Adjacency and Affinity

Independent of any image data, we think of the digital grid system defined by the spels as having a *fuzzy adjacency relation*. This relation assigns to every pair (c, d) of spels a value between zero and one. The closer c and d are spatially to each other, the greater is this number. This is intended to be a "local" phenomenon. How "local" it ought to be should depend on the blurring property of the imaging device. We denote the fuzzy adjacency relation by α and the degree of adjacency assigned to any spels (c, d) by $\mu_\alpha(c, d)$.

Now consider the grid points (spels) as having scene intensities assigned to them. That is, we are given a scene $\mathcal{C} = (C, \mathbf{f})$. We define another local fuzzy relation called *affinity* on spels denoted

by κ. The strength of this relation between any spels c and d, denoted $\mu_\kappa(c,d)$, lies between zero and one, and indicates how the spels "hang together" locally in the scene. $\mu_\kappa(c,d)$ is determined based on $\mu_\alpha(c,d)$, as well as on how similar the intensities or intensity-based properties at c and d are. The functional form of affinities plays an important role in algorithmic efficiency as discussed later on. The properties of fuzzy affinity relations are studied extensively and guidance as to how to setup fuzzy affinities in practical applications is given in [163]. One practical recommendation is *scale-based affinity*. It has been shown that using scale-based affinity leads to significantly improved image segmentation [163].

The idea of "scale" used here, is to determine the size of local structures under a prespecified region-homogeneity criterion. Scale in a scene $\mathcal{C}$ at any voxel $c \in C$ is defined as the radius $r(c)$ of the largest ball centered at c which lies entirely within the same object region (defined under a prespecified region-homogeneity criterion). This adaptive scale concept has evolved in connection with fuzzy connectedness principles [189] and represents the local size of the object at the voxel. In [163], a simple and effective algorithm has been described which estimates $r(c)$ at every location $c \in C$ in any scene $\mathcal{C}$ without explicit object segmentation but based on continuity of region homogeneity. Scale computation and its use in image registration is described in more details in Section 5.2.2.

Generally, both the homogeneity- and object-feature-based components should be considered in the design of fuzzy spel affinities, although in some applications each component on its own may be adequate. The general formulation for affinity is $\mu_\kappa(c,d) = \mu_\alpha(c,d)g(\mu_\psi(c,d),\mu_\phi(c,d))$, where μ_ψ and μ_ϕ represent the homogeneity-based and object-feature-based components of affinity, respectively, and g is a function that combines the two fuzzy values. Both μ_ψ and μ_ϕ are computed using the scale region around the spels [163].

3.2.2 Fuzzy Connectedness and Fuzzy Connected Objects

Our aim is to capture the global phenomenon of "hanging togetherness" in a global fuzzy relation on spels called *fuzzy connectedness*, denoted K. The *strength* of this relation $\mu_{\mathrm{K}}(c,d)$ between any spels c and d, indicating the strength of their connectedness, lies between zero and one, and is determined as follows. There are numerous possible "paths" within the scene domain C between c and d. Each path for our purposes is a sequence of spels, starting from c and ending in d, with the successive spels being nearby. (In general, "nearby" means that the successive spels v and v' are such that $\mu_\alpha(v,v') \neq 0$.) We think of each pair of successive spels as constituting a link and the whole path to be a chain of links. We assign a strength (between zero and one) to every path which is simply the smallest pairwise spel affinity along the path. This indicates the weakest link in the chain. Finally, the strength of connectedness between c and d is the strength associated with the strongest of all paths between c and d.

Let $\theta \subset [0,1]$ denote a set of values of strength of connectedness. (In our practical cases, $\theta = \theta_x = [x,1]$, for some $x \in [0,1]$.) In scene $\mathcal{C}$, a *fuzzy connected object* $\mathcal{O}$ of strength θ, and containing a spel o consists of a pool $O \subset C$ of spels together with a value indicating "objectness" assigned to every spel. Set O is such that $o \in O$, for any spels c and d in O, the strength of connectedness between them $\mu_{\mathrm{K}}(c,d) \in \theta$, and for any spels $c \in O$ and $e \notin O$, the strength $\mu_{\mathrm{K}}(c,e) \notin \theta$. The value of objectness assigned to spels in O varies between zero and one and the value assigned to spels outside O is zero. See [187] for a precise mathematical definition. There are several choices as to how to assign objectness to spels in O, see [163, 187]. A K-*connectivity scene* of a scene $\mathcal{C} = (C, \mathbf{f})$ with respect to a spel $o \in C$ and fuzzy affinity κ in C is a scene $\mathcal{C}_o = (C_o, f_o)$ such that $C_o = C$, and for any spel $c \in C$, $f_o(c) = \mu_{\mathrm{K}}(o,c)$.

3.2.3 Relative Fuzzy Connectedness

Now suppose that scene $\mathcal{C}$ is composed of $m \geq 2$ regions corresponding to m objects $O_1, O_2, \ldots, O_m$, O_m being the background. If $m = 1$, the scene contains information about only one object (no background or only background and no object), and it becomes impossible to perceive the object, and therefore, segmentation of the scene has no meaning. Let $\kappa_1, \kappa_2, \ldots, \kappa_m$ be affinities specified for the m object regions, respectively. These affinities are combined, as described in [160], into a single affinity κ that retains as much as possible the individual object-specific affinity information. Let $o_1, o_2, \ldots, o_m$ be the reference spels selected in objects $O_1, O_2, \ldots, O_m$, respectively. For any $o \in S = \{o_1, o_2, \ldots, o_m\}$, we shall use $b(o)$ to denote $S \setminus \{o\}$, the idea being that $b(o)$ denotes the set of reference spels chosen in the object regions that represent the background regions as far as the object to which o belongs is concerned.

To every path connecting every pair of spels, a strength of connectedness is assigned which is simply the smallest pairwise affinity of spels (as per κ) along the path. Then any spel c is considered to belong to that object with respect to whose reference spel c has the highest strength of connectedness. This relative strength of connectedness, which is modulated as best as possible by considering the affinities that are best for the individual objects, offers a natural mechanism for partitioning spels into regions based on how the spels hang together among themselves relative to others. Since κ has a spatially-variant property that adjusts to the affinities that are best for the individual object regions, a powerful segmentation strategy results that takes into account not only the image properties of all object regions but also their relative spatio-topological arrangement in delineating any given object. *Relative fuzzy connectedness* eliminates the need for selecting a threshold value θ_x on the K-connectivity scene by having the objects "compete" for the spels.

The concepts described above can be extended, as described in [159], to defining fuzzy connected objects when multiple seeds are given for each object. This is the strategy that is employed in most applications mentioned in this thesis.

3.3 Algorithms

To explain our motivation for speeding up the fuzzy connectedness algorithm and for systematically evaluating the new algorithms, an application context becomes necessary. We shall utilize our application relating to the segmentation and quantification of multiple sclerosis (MS) lesions of the brain [164, 189] from dual-echo T2 and proton density (PD) scenes for these purposes. The operator recognition help required in this application is provided in two sessions: (i) by specifying a few seed points for indicating the white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) (and not for lesions), and (ii) by accepting/rejecting each potential fuzzy 3D lesion object detected. After the first session of specifying the seed points, various 3D fuzzy connected objects representing the WM, GM, CSF and the lesions are delineated automatically within the two-component (T2, PD) scene. Nine 3D trackings of fuzzy objects is called for in this method (see [189] for details) including the lesion objects delineated at different strengths. Consequently, in a clinical trial set up, when experimenting with the different parameters of the T2-PD protocol, repeated segmentation experiments involving several patient data sets can take from several hours up to several days between the first and the second interactive session. A speed up of the basic fuzzy connectedness algorithm can substantially facilitate practical applications such as MS lesion analysis.

Generally, in the fuzzy connectedness segmentation method, there are two major computational tasks: (i) computing the fuzzy affinity relations, and (ii) computing the fuzzy connectedness relations. In the original fuzzy connectedness algorithms [187, 189], (i) and (ii) were not separate tasks. In this chapter, we shall refer to (i) as "affinity computation" and (ii) as "tracking" a fuzzy object. Both tasks can be optimized and a dramatic improvement can be achieved in their speed.

Although affinity computation is straightforward, it has to be done separately for each object since the intensity features used for different objects may be different. This step was already op-

timized for speed by using integer arithmetics and lookup tables. However, if we know certain properties of the scenes and the objects in advance, we can minimize affinity computation by avoiding any computations for spels that will definitely not be used during tracking. We shall discuss several strategies along these lines.

The computation of fuzzy connectedness values is accomplished via a graph search algorithm, which may be speeded up if efficient search strategies and data structures are used in the implementation. Most of the results presented here address this efficiency problem. In the rest of this section, we shall first present two groups of algorithms (in Section 3.3.1) and then describe the strategies for speeding up affinity (in Section 3.3.2) and fuzzy connectedness computation (in Section 3.3.3).

3.3.1 Graph Search Algorithms

Computing the fuzzy connectedness values for a fuzzy object is a variation of the single-source-shortest-path problem. This important graph theory problem and the associated algorithms are described elsewhere. See [3, 30, 176] for details and excellent surveys on these topics. We now present two groups of algorithms in a general form so that the different variants, strategies and data structures used, all can be described in the same framework. These are modified versions of well-known algorithms for solving the single-source-shortest-path problem.

Both approaches are iterative. They assign tentative strength labels to spels at each step; the strength labels are estimates of (i.e., lower bounds on) the strengths of the strongest paths from the source (seed spels) to the individual spels. The approaches vary in how they update the strength labels from step to step and how they "converge" toward the strengths of the strongest paths. Label-setting algorithms [3] designate one label as permanent (optimal) at each iteration. In contrast, label-correcting algorithms [3] consider all labels as temporary until the final step, when they all become permanent. The label-correcting algorithms are more general and apply to all classes of problems. The label-setting algorithms are much more efficient, but are applicable only to special situations.

When the strength of object θ_x is specified in advance, the algorithms can make essential use of this information and terminate faster. To keep the presentation simple, the algorithms presented here do not include the details as to how these improvements are actually implemented. These details and the results will be given in Section 3.4.

The original algorithms described in [187] fall in the label-correcting class. This is presented in Algorithm 3. The label-setting group is presented in Algorithm 4. The algorithms of this group are basically various implementations of Dijkstra's well-known algorithm [35].

3.3.2 Affinity Computation

We tested two strategies for computing the fuzzy affinity relations: (S1) pre-compute affinities for all pairs of spels having non-zero adjacency (i.e., neighboring spels in our implementation) before tracking commences, and use as lookup tables during tracking, and (S2) compute affinities on-the-fly, when it is needed by the tracking algorithm for a certain pair of spels (and store the computed values so that they can be reused by simple table lookup if referenced again).

(S1) is a good choice if the threshold for the fuzzy object is not known in advance and we want to experiment with different thresholds to find the optimum for the given fuzzy object. It may also be used if the connectedness values are to be computed for the whole scene (e.g., to be thresholded at different thresholds to extract fuzzy objects of different strengths, as used in our lesion quantification method). However, if a threshold is known in advance (e.g., the final threshold for the object, or the smallest from several thresholds to be used in case of the computation of fuzzy objects of different strengths), computing the affinity values for pairs of spels that are not used in tracking is a waste of time, and so (S2) may be more efficient.

Algorithm 3: Label-correcting

Input: $\mathcal{C}$, o, κ as defined in Section 3.2.

Output: A K-connectivity scene $\mathcal{C}_o = (C_o, f_o)$ of $\mathcal{C}$.

AuxData: An nD array representing $\mathcal{C}_o$ and a queue Q of spels. We refer to the array itself by $\mathcal{C}_o$ for the purpose of the algorithm.

begin

1 set all elements of $\mathcal{C}_o$ to 0 except o which is set to 1;
2 push all spels $c \in C_o$ such that $\mu_\kappa(o, c) > 0$ to Q;
3 **while** Q *is not empty* **do**
4 remove a spel c from Q;
5 find $f_{\max} = \max_{d \in C_o}\{\min(f_o(d), \mu_\kappa(c, d))\}$;
6 **if** $f_{\max} > f_o(c)$ **then**
7 set $f_o(c) = f_{\max}$;
8 push all spels e such that $f_{\max}$, $f_o(e)$, and $\mu_\kappa(c, e)$ satisfy certain conditions (see Section 3.3.3);
endif
endw
end

Algorithm 4: Label-setting

Input: $\mathcal{C}$, o, κ as defined in Section 3.2.

Output: A K-connectivity scene $\mathcal{C}_o = (C_o, f_o)$ of $\mathcal{C}$.

AuxData: An nD array representing $\mathcal{C}_o$ and a queue Q of spels. We refer to the array itself by $\mathcal{C}_o$ for the purpose of the algorithm.

begin

1 set all elements of $\mathcal{C}_o$ to 0 except o which is set to 1;
2 push o to Q;
3 **while** Q *is not empty* **do**
4 remove a spel c from Q for which $f_o(c)$ is maximal;
5 **for** *each spel* e *such that* $\mu_\kappa(c, e) > 0$ **do**
6 set $f_{\min} = \min\{f_o(c), \mu_\kappa(c, e)\}$;
7 **if** $f_{\min} > f_o(e)$ **then**
8 set $f_o(e) = f_{\min}$;
9 **if** e *is already in* Q **then**
10 update the location of e in Q;
else
11 push e to Q;
endif
endif
endfor
endw
end

The details of affinity computation are not given here. They directly follow from the functional forms of the affinity relations. The following functional forms were used for the fuzzy affinity relations in all our experiments. We consider the given vector-valued scene in the MRI case to be equivalent to two scenes $\mathcal{C}_{\mathrm{PD}} = (C, f_{\mathrm{PD}})$ and $\mathcal{C}_{\mathrm{T2}} = (C, f_{\mathrm{T2}})$.

$$f_{\mathrm{ave}}(a,b) = \frac{1}{2}(a+b)\,, \tag{3.1}$$

$$f_{\mathrm{reldiff}}(a,b) = \begin{cases} \frac{|a-b|}{a+b}, & \text{if } a+b \neq 0, \\ 0, & \text{otherwise,} \end{cases} \tag{3.2}$$

$$G(x,m,\sigma) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} e^{-\frac{1}{2\sigma^2}(x-m)^2}\,, \tag{3.3}$$

$$g_{pq}(c,d) = G(f_q(f_p(c), f_p(d)), m_{pq}, \sigma_{pq})\,, \tag{3.4}$$

$$g_p(c,d) = \min_q g_{pq}(c,d)\,, \tag{3.5}$$

$$\mu_\kappa(c,d) = \mu_\alpha(c,d)\left((g_{\mathrm{PD}}(c,d))^{w_1}(g_{\mathrm{T2}}(c,d))^{w_2}\right)\,. \tag{3.6}$$

In Eq. (3.3), G is a Gaussian function of x with mean m and standard deviation σ. In Eq. (3.4), G is used with actual mean and standard deviation parameters m_{pq} and σ_{pq}, where p and q are indexes such that $p \in \{\mathrm{PD}, \mathrm{T2}\}$, and $q \in \{\mathrm{ave}, \mathrm{reldiff}\}$. In Eq. (3.6), w_1 and w_2 are nonnegative weights such that $w_1 + w_2 = 1$. In our example application, we use $w_1 = 2/3$ and $w_2 = 1/3$ for WM and GM objects, and $w_1 = 1/3$ and $w_2 = 2/3$ for CSF. For computational simplicity, we used the six-adjacency relation for α, weighted by the appropriate voxel dimensions. That is,

$$\mu_\alpha(c,d) = \begin{cases} 1, & \text{if } c = d, \\ \frac{\min(s, \|c-d\|)}{\|c-d\|}, & \text{if } c \text{ and } d \text{ are 6-adjacent,} \\ 0, & \text{otherwise,} \end{cases} \tag{3.7}$$

where s represents the length of a side of the square pixels and $\|c-d\|$ represent Euclidean distance between the centers of c and d.

3.3.3 Tracking

Although there is no single "best" strategy (different strategies may be the "best" for different kinds of applications, and different affinity relations), we found a few variants that are good approximations to the (possibly) best in most cases. We now present the details of the different strategies and data structures used for tracking. We assigned codes for the methods according to a four-digit numbering scheme wherein each digit represents some property of the method/data structure used. These codes will be used throughout Section 3.4.

The different data structures we utilized include: queue, priority queue, various heaps, LIFO and FIFO lists, hash tables with different hashing functions and table sizes. The strategies used are: different criteria for putting a spel in a queue, using different data structures to maintain the priority queue, using different hash functions (i.e., hashing by geometry-based properties or by affinity-based properties), and strategies for addressing the memory. A discussion of these data structures is beyond the scope of this thesis. See [3, 30, 176] or any standard textbook on data structures for a detailed discussion.

Label-correcting Algorithms

The condition used in Step 8 of Algorithm 3 in the original fuzzy connectedness algorithm [187] was

$$\mu_\kappa(c,e) > 0\,. \tag{3.8}$$

This means that whenever the label of a spel c is updated, all neighboring spels e (with positive affinity) are pushed to the queue Q, because the strongest-so-far path to e might be improved through c. This resulted in many unnecessary push operations (and therefore unnecessary pop operations later), since, in most cases, a change in a spel label has effect only on a few neighbors.

The implementation used in [189] achieved a big practical improvement by replacing the condition in Step 8 of Algorithm 3 by

$$f_{\max} > f_o(e)\,. \tag{3.9}$$

Carefully examining all possible relationships between $f_{\max}$, $f_o(e)$ and $\mu_\kappa(c,e)$, we established a theoretically optimal condition for Step 8 of Algorithm 3:

$$f_{\max} > f_o(e) \quad \text{and} \quad \mu_\kappa(c,e) \geq f_o(e)\,. \tag{3.10}$$

Since it means evaluating a more complicated expression, in real applications, its use may not really pay off. This may be the case, if additional techniques are involved to improve the tracking speed, such as an additional bit-array to keep track of whether a spel e is already in Q, in which case it need not be duplicated. This latter addition can be used with any of the above three conditions in Eqs. (3.8)–(3.10). We use codes 00b0, 00b1, and 00b2 for the algorithms using condition in (3.8), (3.9), and (3.10), respectively, where b is either 0 (no additional bit-array) or 1 (use additional bit-array).

Label-setting Algorithms Using *d*-ary Heap

Finding elements other than the one with the maximum key (i.e., the root) in a heap is usually not really efficient unless some other helper structures are used for the search. Our versions of the tracking algorithm using heaps use different helper structures and have different efficiencies.

We represent the priority queue Q in Algorithm 4 by a d-ary heap. Here, d refers to the maximum number of 'children' any node in the heap may have. Since this notation is in standard use in the literature, we keep it for the rest of this chapter. However, d here is not to be confused with d used to denote spels in other parts of the thesis. We use a binary heap ($d = 2$) unless noted otherwise. The key of a spel c in Q is the tentative strength label $f_o(c)$ of c at the time it is inserted into Q. Since the root stores the element with the largest key, Step 4 is the combination of a *find-max* and a *delete-max* operation. The push operation in Steps 2 and 11 is the *insert* heap operation, and the update in Step 10 is the *increase-key* operation.

In the first version of d-ary heap (denoted by 1000), we do not keep track of whether an element is already in the heap, and we do not perform a search in the heap for an already stored element, so we always insert a new instance of e in Steps 10 and 11, even if it means duplication. In another version (denoted by 1010) we maintain an additional pointer array, which, for every voxel v stores the position of v in the heap (or a marker that v is not in the heap). This array is used in the test in Step 9 and by the update (*increase-key*) operation in Step 10. We shall come back to the memory versus speed issue later.

The tracking algorithms in our final group using d-ary heaps are more memory-conserving. They use hash tables of various sizes with various hash functions to keep track of the positions of the voxels in the heap to make the *increase-key* operation used in the update (Step 10) efficient. However, maintaining these hash tables requires additional computation, and different hash functions work differently depending on the distribution of the key values assigned to the input data (i.e., depending on the affinity values and on the geometric structure of the fuzzy object tracked).

The four hash functions we implemented assign a key value to a heap element c (a spel, in our

case) using the following formulas:

$$\text{key}(c) \equiv ((c_3 * \text{height} + c_2) * \text{width} + c_1) \pmod{\text{TS}}, \tag{3.11}$$

$$\text{key}(c) \equiv (c_3 + c_2 + c_1) \pmod{\text{TS}}, \tag{3.12}$$

$$\text{key}(c) \equiv (c_3 * c_2 * c_1) \pmod{\text{TS}}, \tag{3.13}$$

$$\text{key}(c) \equiv (c_3 \text{ XOR } c_2 \text{ XOR } c_1) \pmod{\text{TS}}, \tag{3.14}$$

where c_1, c_2, c_3 are the coordinates of the spel (voxel) c, height and width are the dimensions of a slice of the scene, TS is the size of the hash table, and XOR is the bit-wise exclusive or operation. The first hash function is the commonly used linear addressing scheme of elements in multidimensional arrays. The remaining represent some simple ways of combining the three coordinates into a single key value. TS depends on the actual hash function and was set up by studying the distribution of the key values with different values of tablesize. For Eqs. (3.12) and (3.14), we used $\text{TS} = 768$ and $\text{TS} = 256$, respectively, that creates a separate hash bin for each possible combination of 8-bit coordinates (i.e., the modulo operator never merges two distinct values). For Eqs. (3.11) and (3.13), the range of possible values is quite large and hashing is really necessary. We selected $\text{TS} = 8191$ for these cases, since this is reasonably small yet it results in a fairly uniform distribution of hash keys in our application (hence efficient hashing). We use the codes 1020, 1021, 1022, and 1023, respectively, for the algorithms corresponding to Eqs. (3.11)–(3.14).

Label-setting Algorithms Using Fibonacci Heap

The Fibonacci heap [30, 49] is a data structure that allows the heap operations to be performed more efficiently than *d*-ary heaps. We used three versions of the tracking algorithm, analoguous to those encoded 1000, 1010, 1020, and the corresponding codes differ only in the second digit: 1100, 1110, 1120.

With both types of heap structures, when hash tables are used to help the search for arbitrary elements in the heap, the hash function assigns the key value based on the coordinates of the spel, which we may call hashing based on geometry.

Label-setting Algorithms Using Dial's Implementation

Dial [34] gave a practical, very efficient implementation of Dijkstra's shortest path algorithm. We modified his data structure and algorithm to compute fuzzy connectedness values. In our problem, since each voxel's label (connectedness value) is bounded by the maximum possible affinity value (which in our cases is at the most 65535), we can use the temporary connectedness strength label directly as a hashing key. We maintain a list of buckets, numbered $0, 1, \ldots, A$, where A is the maximum possible (scaled integer) affinity value in the application, and each bucket k stores all nodes with the temporary connectedness value equal to k. The nodes within the same bucket are stored in a doubly linked list, so that both LIFO and FIFO type of operations can be performed on it. When a spel c is entered into the priority queue Q (*insert*), its actual temporary label $f_o(c)$ is used as a hashing key, and c is put into the end of the corresponding bucket's list.

We tried two versions of the above described algorithm: with the LIFO version of *find-max* (denoted by 1200), and with the FIFO version (denoted by 1201). We tested also a version, where an additional pointer array is used (as with the heaps) to make direct access of elements in the queue possible without searching through the bucket lists. This latter version is referred to by the code 1210. For ease of reference, and for comparison, Table 3.1 summarizes the various methods described so far and the associated codes.

Table 3.1: Summary of the various methods used for comparison and their associated codes.

code	method[a]	data structure	pointer array	hashing[b]	condition
0000	L-C	queue			Eq. (3.8)
0001	L-C	queue			Eq. (3.9)
0002	L-C	queue			Eq. (3.10)
0010	L-C	queue	yes		Eq. (3.8)
0011	L-C	queue	yes		Eq. (3.9)
0012	L-C	queue	yes		Eq. (3.10)
1000	L-S	d-ary heap			
1010	L-S	d-ary heap	yes		
1020	L-S	d-ary heap		Eq. (3.11)	
1021	L-S	d-ary heap		Eq. (3.12)	
1022	L-S	d-ary heap		Eq. (3.13)	
1023	L-S	d-ary heap		Eq. (3.14)	
1100	L-S	Fibonacci heap			
1110	L-S	Fibonacci heap	yes		
1120	L-S	Fibonacci heap		Eq. (3.11)	
1200	L-S	Dial's / LIFO		connectedness[c]	
1201	L-S	Dial's / FIFO		connectedness[c]	
1210	L-S	Dial's / FIFO	yes	connectedness[c]	

[a] L-C: label-correcting; L-S: label-setting.
[b] All hashing functions include a modulo TS (table size).
[c] Direct hashing by the connectedness value.

3.4 Experiments, Results, and Discussion

The presented algorithms were compared using images of different protocols. Image data is described first and methods of comparison are given next.

3.4.1 Image Data and Task

For the results presented in this paper, we utilized 10 MR image data sets acquired with a fast spin-echo, dual echo protocol (with two intensity values per voxel) (47 to 55 slices of size 256×256 per volume). The task was to classify the component tissues (white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF)) of the brain. This is a common step required in many neurological applications. All data sets were first standardized by our MR intensity standardization method [128, 135] to make them have well defined intensity scale for the main foreground objects (i.e., the brain tissues, in our case).

The parameters for the affinity relations for the different fuzzy objects (tissues) were estimated from a separate set of pre-segmented datasets of the same protocol that underwent the same intensity standardization. The segmentation of these training sets was verified and manually corrected by a radiologist when found necessary. Figure 3.1 shows a central slice of a standardized MRI study pair (PD, T2, first row), the corresponding slices from the scenes depicting the fuzzy affinity relations for the three objects (second row), the same slices from the respective connectivity scenes depicting the connectedness values (third row), and the hard (binary) objects thresholded from the connectedness values (last row).

We also utilized a CT scene pertaining to a craniofacial study (97 slices of size 512×512) and an

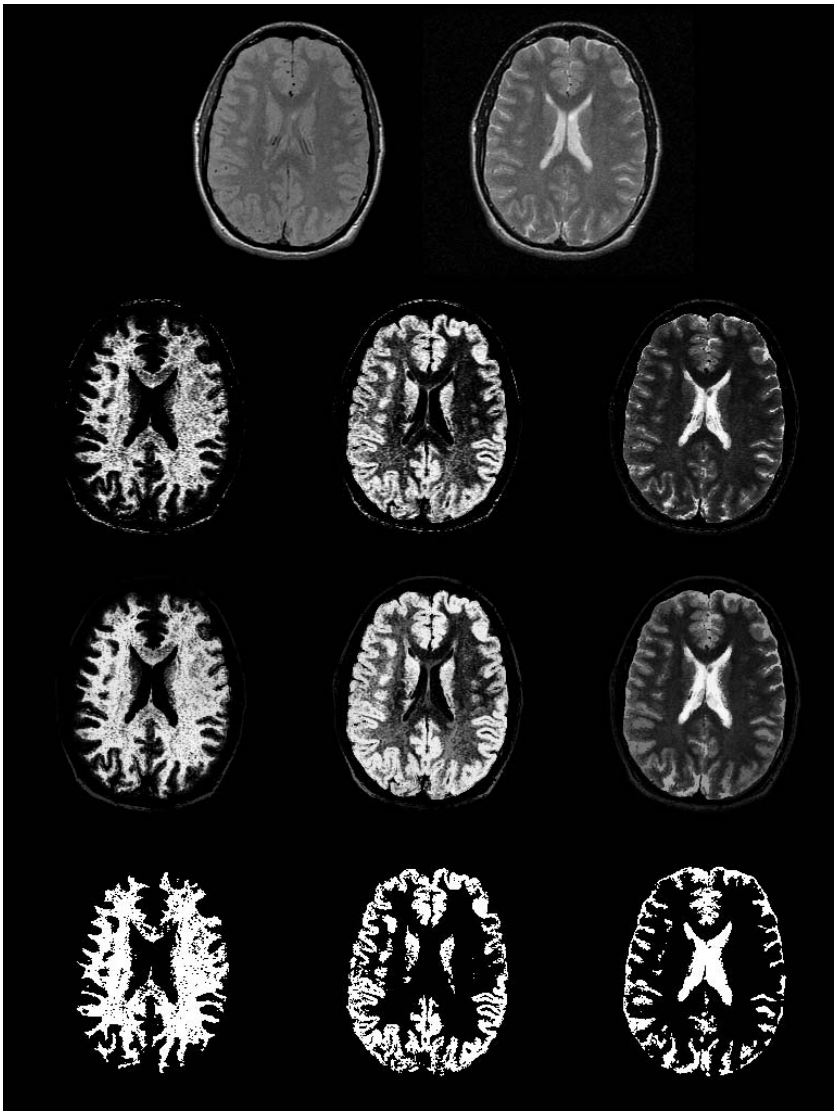


Figure 3.1: Slices from a standardized MR study pair (PD, T2, first row), the corresponding slices from the scenes depicting the fuzzy affinity relations for the three objects (WM, GM, CSF, second row), the same slices from the scenes depicting the connectedness values (third row), and the hard (binary) objects thresholded from the connectedness values (last row).

MR angiographic (MRA) scene with a contrast agent (60 slices of size 512×512) in order to evaluate the different tracking strategies on completely different input scenes, affinities, and objects. Slices of the original images, images depicting the affinity values, the connectedness values, and the final hard object are shown in Figs. 3.2 and 3.3.

We wish to emphasize here that the segmentation of these data sets is challenging, practically relevant and therefore does not represent an artificial example. For the CT case, effective segmentation of the bony structures that include thin and softer bones in addition to the easier hard cortical bones but that separates soft tissues (such as muscles and skin) which have CT properties similar to thin and cancellous bones is quite difficult. Fuzzy connectedness has been demonstrated to provide

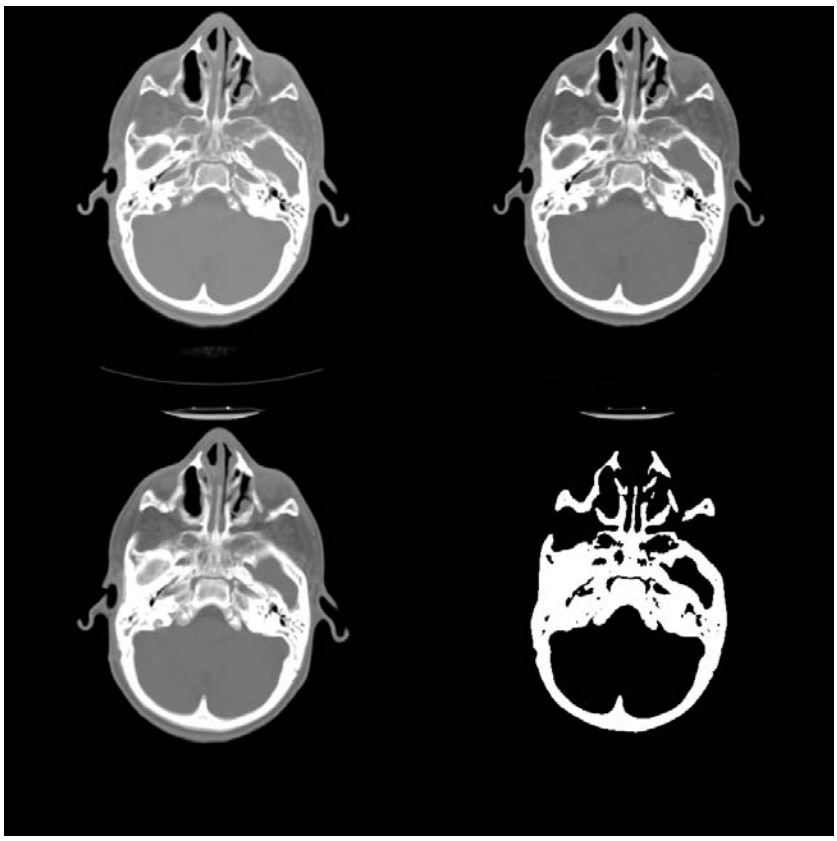


Figure 3.2: A slice of a CT dataset (top left), the corresponding slices of the scene depicting the affinity values (top right), the connectedness values (bottom left), and the final hard object (bottom right).

a tangible solution in this case [182]. For the MRA case, the delineation of big and finer vessels simultaneously poses a similar challenge for which fuzzy connectedness has been demonstrated to be effective [94, 155].

Since the graph corresponding to our problem is quite regular and sparse, the algorithms that have better performance on sparse graphs gave better results. However, due to the nature of the affinity relations and the fuzzy objects in our applications, some of the theoretically not-so-efficient algorithms considerably outperform theoretically optimal algorithms.

The average number of voxels per scene was 3,342,336 for the 10 MRI scenes. The average segmented volume in voxels ($\pm$ standard deviation) was 217,381 ($\pm$26,657) for WM, 260,574 ($\pm$23,398) for GM, and 167,171 ($\pm$31,666) for CSF, that is 6.50% ($\pm$0.71%), 7.81% ($\pm$0.77%), and 4.99% ($\pm$0.82%), respectively, of the scene domain. Among the three objects, WM is the simplest in shape, GM shares a complex surface with the peripheral CSF, and CSF also shares another similar surface with the dura (considered as background in this application). The peripheral CSF has very fine structures, and the partial volume effects are most severe on the CSF-background and CSF-GM boundaries. Due to the differences in these structural shapes, the actual tracking time is not proportional to the volume of the fuzzy objects.

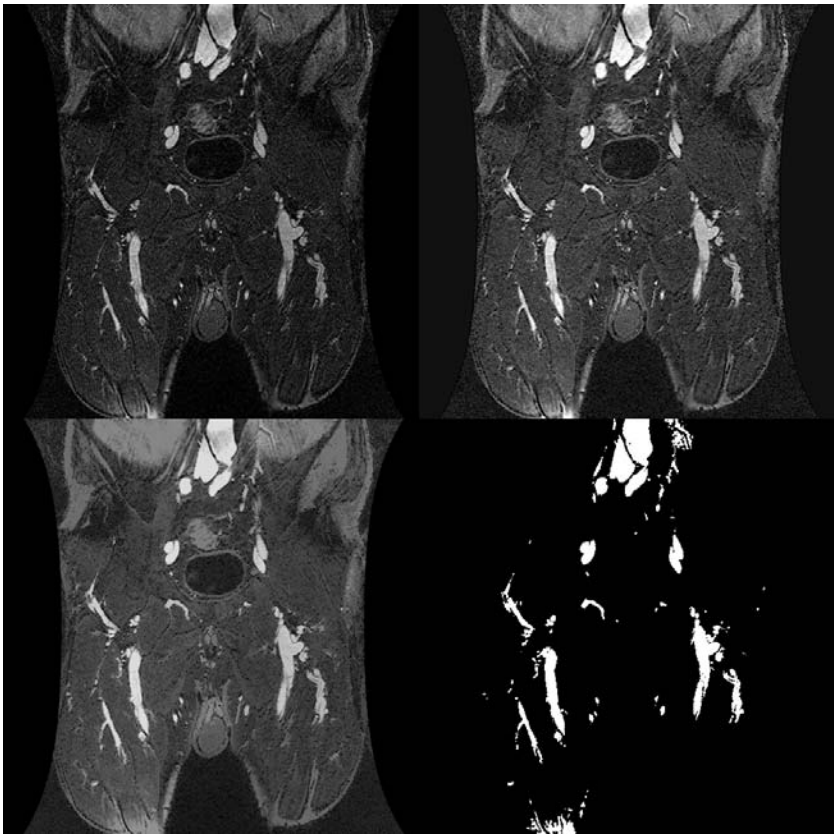


Figure 3.3: A slice of an MRA dataset (top left), the corresponding slices of the scene depicting the affinity values (top right), the connectedness values (bottom left), and the final hard object (bottom right).

3.4.2 Comparisons

In this section, we compare the performance of the 18 implemented algorithms in different hardware and software environments, with different parameter settings on several types of image data.

Hardware

Our main observation is that, on average, the PC with a 1.4 GHz AthlonXP CPU running Linux 2.4 is approximately 12 times faster than a Sun SPARCstation-20 with a 50 MHz CPU running Solaris 2.5, when using the fastest algorithms (1200, discussed later), about 16 times faster when using close-to-best algorithms (1110 and 1210), and 28–36 times faster with the slower label-correcting algorithms. It is clear that the speed ratio is not directly proportional to the difference in the CPU clock frequencies. The speed factor between the two hardware platforms is more when using the label-correcting algorithms that use about 20–30 times more iterations than the label-setting algorithms. There is something happening behind the scenes (perhaps having to do with the CPU cache size/speed and/or bus speed) but we have not explored this aspect. It is merely an observation in our case by simply looking at the figures. For the rest of this paper, all results shown were measured on the PC platform.

Table 3.2: Average running time (in seconds) over 10 MRI scenes for fuzzy object tracking including affinity computation. Comparison of the cases of pre-computed affinities and on-the-fly computation for several objects.

how	what to compute	WM	GM	CSF
step-by-step	Affinity + Connectivity	22.12	23.87	25.60
	Affinity only	16.91	15.50	15.59
on-the-fly	Connectivity(+Affinity) whole	19.00	26.27	32.44
	Connectivity(+Affinity) with threshold	5.65	6.87	3.53

Affinity Computation

In this comparison, since our aim was to measure the speed improvement that can be gained if the affinity computation can be restricted in some way, we used one of our fastest tracking algorithms (1210) to do the actual tracking in each case.

The results in Table 3.2 show that when affinity is computed on-the-fly, tracking time is shorter for WM and longer for GM and CSF, as compared to the step-by-step method. Since the affinity values for the background spels are not computed, it would save about 75% of computation. However, when processing on-the-fly, additional function calls for each referred affinity value create some overhead. This overhead is comparable to the saving on this AthlonXP machine, therefore, the real savings are small (10% for WM) or in some cases, there is not any (GM and CSF). Further, the on-the-fly algorithms can make essential use of pre-determined thresholds, and the entire computation may take less than 30% for WM and GM, and 10% for CSF, as compared to that without thresholds.

Although the speed is still not really interactive, if affinity needs to be computed during the interactive process, in many cases, it is sufficient to do the computation and the tracking only on a 2D slice. This can be done in real time, allowing interactive parameter settings for the affinities and threshold selection. The connectivity scene (and the final object segmentation) can then be obtained for the whole 3D (or even 4D) scene in just a few seconds.

In the following sections, we present a comparison of the various tracking algorithms listed in Table 3.1. All results include only tracking time. That is, the fuzzy affinities were pre-computed for the whole volume before the tracking (and time measurement) took place.

Tracking Without Preset Threshold

The left table in Table 3.3 shows the average tracking time (over the 10 MRI scenes) needed for the three objects using the different tracking algorithms. The algorithms are sorted in "goodness" order from top to bottom, the best appearing at the bottom.

Our first observation is that the label-setting algorithms perform much better than the label-correcting variants, as we expected. The two exceptions are 1023 and 1021, where the choice of the hash function does not seem to be very useful. These two hashing versions have very small hash tables, hence long lists for each bin. However, the worst label-setting algorithm (besides the two extremes mentioned above) is still more than two times faster than the best label-correcting algorithm. Further, there is a factor of almost three difference in the speed between the best and the worst label-setting algorithms.

As for the label-correcting versions (00xx), those using the optimal condition in Step 8 of Algorithm 3 (0002) are 20–25% faster than those with the condition in the previous (extensively used) implementation (0001) [187, 189], and 15–20% faster if using the binary map helper structure to avoid multiple instances of the same voxel in the queue (0012 and 0011). The additional memory

Table 3.3: Average tracking time (in seconds) over 10 MRI scenes for different algorithms for different fuzzy objects using pre-computed affinities without using pre-determined thresholds for the objects (left) and with thresholds (right).

algorithm	WM	GM	CSF
1023	35.47	71.82	109.48
1021	30.71	61.15	90.19
0000	22.19	48.28	70.05
0001	19.35	38.89	53.23
0002	15.40	29.97	40.84
0011	14.33	27.63	35.29
0010	13.18	27.42	36.42
0012	13.20	25.11	32.24
1022	5.94	10.46	16.51
1201	5.98	7.92	18.52
1120	5.95	9.27	12.42
1020	5.48	8.69	11.70
1100	5.60	8.20	10.36
1010	4.94	7.12	9.02
1000	4.64	7.00	8.56
1110	4.46	6.19	8.11
1210	2.56	4.17	4.93
1200	2.17	2.95	6.15

algorithm	WM	GM	CSF
0001	8.25	14.09	5.25
0000	8.43	13.27	4.66
0002	6.55	11.44	5.36
0011	6.13	10.67	4.81
0012	5.71	9.77	4.66
0010	5.02	7.99	3.39
1023	3.43	4.05	0.96
1021	3.15	3.64	0.92
1100	1.51	1.69	0.65
1000	1.26	1.51	0.53
1010	1.23	1.49	0.54
1120	1.28	1.45	0.54
1022	1.19	1.44	0.55
1020	1.14	1.39	0.54
1110	1.09	1.22	0.50
1210	0.60	1.06	0.23
1201	0.62	0.89	0.21
1200	0.38	0.50	0.18

needed for the binary maps ($<$.5 MB) is not really significant compared to the rest of the scene data we need to store (altogether $\approx$39 MB).

The use of Fibonacci heaps (11xx) compared to binary heaps (10xx) seems to be slightly worse (mainly due to the more complicated data structure management and the sparseness and very simple structure of the graph in case), although in case of the additional direct pointers (1110 and 1010), the Fibonacci heap is faster. However, the memory storage required for these direct pointers is quite large (13 MB for the size of our scenes)—one third of the total other memory required—and there is no speed increase with the binary heap and that is approximately 10% with the Fibonacci heap.

The algorithms based on Dial's implementation 12xx proved to be the best for computing all connectedness values. Using 1200 and pre-computed affinities, the tracking time for the connectedness values for WM, GM, and CSF objects is 2.17, 2.95, and 6.15 seconds, respectively, on average. This is certainly interactive time and is about half the time taken by the best heap algorithm to perform the same task. The FIFO version seems to perform much better that the LIFO version; it is twice as fast for WM and GM. However, for CSF, it its more than three times faster than the LIFO version. This is due to the more complex and fine structure of particularly the peripheral CSF.

As mentioned earlier, in many cases it is sufficient to do the tracking on 2D slices, which can be done in real time, and after interactive parameter settings, compute all connectedness values for the 3D (or 4D) scene in almost real time. This kind of application setup (i.e., rough 2D then fine 3D) is especially useful when multiple objects need to be segmented from the same data and relative connectedness or iterative/interactive relative connectedness methods are used in segmentation [160] based on the relative strength of connectedness of spels in different objects.

Table 3.4: Average tracking time (in seconds) over the 10 MRI scenes for three algorithms using d-ary heaps, for the three objects, for several values of d.

object	algorithm	$d = 2$	$d = 3$	$d = 4$	$d = 5$
WM	1000	4.640	4.212	4.157	4.226
WM	1010	4.942	4.174	3.996	3.962
WM	1020	5.483	4.983	4.909	4.931
GM	1000	7.004	6.327	6.247	6.345
GM	1010	7.122	5.981	5.745	5.705
GM	1020	8.693	7.951	7.838	7.879
CSF	1000	8.655	7.904	7.815	7.962
CSF	1010	9.120	7.872	7.424	7.399
CSF	1020	11.696	10.797	10.666	10.720

Tracking With Preset Object Threshold

The right table in Table 3.3 shows a comparison of the algorithms in a way similar to that in the previous section except that, in this case, the algorithms made essential use of pre-determined thresholds for the fuzzy objects. The following changes are required to incorporate threshold into the algorithms. In Algorithm 3, Steps 4-8 should be executed only if $f_o(c) < x$, where x is the predetermined threshold (see Algorithm $\kappa\theta_x FOE$ in [187]). In Algorithm 4, the while loop should stop after Step 4 when $f_o(c) < x$ for the first time. The thresholds (0.5 for WM and CSF, and 0.7 for GM, on the [0,1] scale) were arrived at by averaging manually selected thresholds on the connectedness values for a few scenes.

As in Table 3.3(left), the drops corresponding to the label-setting algorithms (1xxx) and the Dial's algorithms (12xx), are significant (especially for CSF). Note also that the two hashing versions (1023 and 1021), that were extremely slow in the whole connectedness computation, are performing significantly better when the thresholds are known in advance, although they are still the worst among the label-setting algorithms. Surprisingly, the two algorithms using quite different hash functions (1020 and 1022) have almost the same performance when only the voxels belonging to the fuzzy objects get inserted into the heap. This was not true without using the thresholds. The few changes in the order among the other techniques are mainly due to the fact that fewer voxels get inserted into the queues/heaps/lists, and therefore, the search for the elements is more effective, and the cost of maintaining the data structure and the cost spent in the tracking loop cover different fractions of the total cost. Here too, the best algorithms are those using affinity-based hashing (12xx), with a performance about 2.5 times better than that of the best heap-version.

Tracking Time for Different *d*-ary Heaps

We also compared the algorithms using d-ary heaps for different values of d. We tested the "stupid" (1000), the "memory-expensive" (1010), and the best of the four hashing versions of the algorithm (1020), with $d = 2, 3, 4, 5$. Table 3.4 shows how the tracking time varies with the change of the d value, for several objects.

According to theoretical complexity analysis [3], the optimal choice would be $d = \max\{2, \lceil m/n \rceil\}$, where n is the number of nodes in the graph and m is the number of arcs. In our application, the nodes are the voxels, and each voxel face induces one arc (since our graph is undirected), hence $m/n = 3$, and the optimal value would be $d = 3$. In a few test runs (using $d = 2, 3, \ldots, 9$), since the measured running time increased starting with $d = 6$, we decided to include only $d = 2, 3, 4, 5$ in the

Table 3.5: Tracking time (in seconds) for different algorithms for the bone fuzzy object in a CT scene and for the blood vessel fuzzy object in an MRA scene, using pre-computed affinities without using pre-determined thresholds for the objects (left) and with thresholds (right).

algorithm	vessel	algorithm	bone
1201	*	1201	*
1023	*	1023	3029
1200	10173	1021	1729
1021	5526	1200	1059
1022	2034	0000	240
1120	679	1022	230
1020	650	1020	179
0000	645	1120	166
0001	421	0001	161
0010	343	0002	136
0002	337	0010	127
0011	285	0011	108
1210	281	0012	101
0012	250	1210	88
1100	107	1010	54
1000	105	1100	51
1010	103	1000	48
1110	85	1110	45

algorithm	vessel	algorithm	bone
1201	434.25	0000	9.12
1023	186.80	0001	7.12
1200	178.95	0002	6.26
1021	96.93	0011	5.25
0000	42.98	0012	4.96
0001	34.10	0010	4.93
0002	29.34	1201	4.36
0010	24.58	1023	2.15
0011	24.44	1021	1.70
0012	23.07	1100	1.23
1022	21.41	1120	1.15
1120	20.59	1010	1.14
1020	19.90	1000	1.11
1100	13.73	1022	1.09
1010	13.38	1020	1.06
1000	12.40	1110	1.02
1110	11.42	1200	0.77
1210	9.08	1210	0.56

* have not finished within 7.5 hours!!!

final experiment. The performance gain using $d = 5$ as compared to $d = 2$ is approximately 10% for algorithms 1000 and 1020, and 20% for algorithm 1010. Our practical optimum for d-ary heap algorithms proved to be $d = 4$ for algorithms 1000 and 1020, and $d = 5$ for algorithm 1010. This points at the need for a practical evaluation of algorithms in addition to their theoretical analysis. The theoretical analysis in [3] does not consider the special structure of the graph representing fuzzy affinity relations, and the distribution of the arc weights (fuzzy affinity values), which are certainly special here, since we are dealing with images, where neighboring voxels have highly correlated values.

Other Data

We compared the algorithms using two additional scenes, very different from the MRI brain scenes illustrated so far, in order to study the performance of the different strategies on different affinity setup and object structures. Tables, similar to those in the previous section, are shown in Table 3.5 for the two objects—bone and blood vessels—in these scenes. Some of the main observations made for the case of MRI brain studies are still valid, although there are significant differences. Without making use of a pre-determined threshold, the affinity-based hashing algorithms with search in the bucket lists (120x) are among the worst of all, since the distribution of the affinity values in this setup is quite far from uniform (i.e., all bone-like spel pairs have very high affinity and everything else is almost zero). However, algorithms using heaps (1110, 1000, 1010 and 1100) seem to perform better on these kinds of data. Method 1210 (one of the bests for the brain data), using the additional pointer-array (which is quite large for such large scenes) is still very competitive requiring about twice as long as the worst of the best heap-based methods for the bone object and about three times as long for the vessel object.

Table 3.6: Comparing the average tracking time spared when using different pre-determined values of threshold x for the fuzzy objects. The time spared is expressed in percent of the time needed for the complete tracking without a threshold.

object	$x = 0.1$	$x = 0.2$	$x = 0.3$	$x = 0.4$	$x = 0.5$	$x = 0.6$	$x = 0.7$
WM	38.36%	51.47%	61.41%	69.39%	76.67%		
GM	24.27%	37.13%	47.97%	58.17%	67.33%	76.16%	84.46%
CSF	60.12%	78.66%	86.77%	91.67%	94.83%		
vessel	23.90%	48.72%	91.13%	97.55%	99.45%	99.80%	
bone	0.00%	10.87%	16.63%	25.89%	89.64%		

However, when the algorithm can make use of pre-determined thresholds, the heap-based algorithms come down to about 15 seconds for the vessel and to about one second for the bone data, and the best (although memory-expensive) 1210 finishes in 9 seconds for the MRA and in half a second for the CT data (see Table 3.5(right)). Whether the memory-expensive method that runs faster or the memory-conserving method that is slightly slower should be used depends on the actual bounds for memory-usage and running-time. This is close to interactive speed and if we consider the application setup utilizing 2D slices in the interactive phase as described previously, the faster algorithms can indeed be used in an interactive environment.

Saving by Using Pre-determined Thresholds

Table 3.6 presents the tracking time spared when using some pre-determined threshold value x, as a function of x, as compared to the case with no preset threshold (i.e., $x = 0$). It is well illustrated here that even a small value of x (such as 0.1, which is significantly less than the final threshold for the object) can save considerable amount of computation (25–60% for the brain tissue objects). It is also shown that the gain is more dramatic in the CT and MRA case than that for the MR brain images. This is due to the different distribution of the affinity values in these cases. Data for one of the fastest tracking algorithms (1210) were used in this comparison. For inferior tracking algorithms, the effect of the small initial threshold is even more substantial (not demonstrated here).

3.5 Concluding Remarks

We presented several algorithms for fuzzy-connected object segmentation in volume images. We presented two groups of algorithms—label-correcting and label-setting—and a variety of strategies of implementation under each group. We tested these algorithms in real medical applications, namely the separation of brain tissues in MRI, the segmentation of the skull in CT, and the segmentation of the blood vessels in MRA images. Existing implementation executes 12–36 times faster on a PC with a 1.4 GHz AthlonXP CPU running Linux 2.4 compared to a 50 MHz Sun SPARCstation-20 running Solaris 2.5. However, this does not offer interactive operations. Utilizing efficient algorithms and careful selection of implementations can speed up the computation of fuzzy connectedness values by a factor of 16 to 29 (on the same hardware), as compared to the implementation previously used in our applications utilizing fuzzy object segmentation. The running time is reduced considerably (by a factor up to 34 for brain MR and even more for bone CT), when the algorithms make use of pre-determined thresholds for the fuzzy objects. In real applications, low thresholds can be used in advance without compromising the final outcome of the segmentation. This can save valuable seconds for real-time, interactive applications. In routine applications (such as MS lesion analy-

sis [189]), the thresholds are known and hence can be preset, resulting in a tremendous saving in time.

We may conclude that there is no "always optimal" algorithm. The optimality of an algorithm depends on the input data as well as on the type of affinity relations used. However, some of the algorithms have definitely consistently better performance than others. The label-setting algorithms are faster in real applications, as the theoretical complexity analysis suggested. However, those using more complicated data structures (such as Fibonacci heaps rather than d-ary heaps) are not the fastest in practice. This is due to the big differences in the "constant" in their theoretical complexity formulas. Many analyses do not account for search for stored elements in queues/heaps/lists that can be eliminated only by large pointer-arrays, which may not be affordable (e.g., $\approx$ 100 MB additional memory in case of the CT scene). In the case of brain MRI studies and brain tissue fuzzy objects, the affinity values are better distributed than those in the CT case. Hence the affinity-based hashing (method 1200) proved to be better for the MRI case, while geometry-based hashing (method 1000) proved better in the CT case when not using additional pointers and searching in lists/buckets. We conclude that it is useful to offer several algorithms and versions in the actual implementation and let the user select the most appropriate for the actual data and task at hand. The following general guideline seems appropriate in selecting an algorithm. For blob-like objects with a spread-out distribution of their intensities or other properties (such as brain tissues in MRI), label setting algorithms with affinity-based hashing are to be selected. For more sparse objects (vessels, bone) with high contrast (MRA, CT), label setting algorithms with geometric hashing are to be chosen. In our implementation in 3DVIEWNIX both strategies are made available. In any large application invoking the processing of 100's of data sets, it is a good idea to spend initially some time on testing which strategy is the most efficient. This will save a substantial amount of time during the life of the application at hand.

When these algorithms were first introduced they could not be used for interactive processing. Today, using the fast algorithms proposed herein on fast (but not necessarily top-of-the-line) hardware, interactive speed (less than one second per scene) of fuzzy object segmentation is achievable. This leads to a new line of segmentation techniques. Utilizing the robust fuzzy connectedness principles and fast algorithms, interactive affinity parameter, seed point, and threshold selection becomes possible and the segmentation can be more efficient (in terms of both the user interaction time and computing time) without compromising the proven precision and accuracy of these techniques. If images having intensities with tissue-specific meaning (such as CT or standardized MR images) are utilized, most of the parameters for the segmentation method can be fixed, fuzzy affinities and other intermediate data can be computed before the user interaction is needed and the user can be provided with more information at the time of interaction. More importantly, for very difficult segmentation tasks, perhaps user-steered techniques can be devised wherein fuzzy connected segments are collected within a 2D/3D/4D region of interest (ROI) swept by the user interactively within the scene domain.

Chapter 4

An MRI Protocol-independent Brain Segmentation Method

We present a segmentation method that combines the robust, accurate, and efficient techniques of fuzzy connectedness with standardized MRI intensities and fast algorithms. The result is a general segmentation framework that more efficiently utilizes the user input (for recognition) and the power of computer (for delineation). This same method has been applied to segment brain tissues from a variety of MRI protocols. Images were corrected for inhomogeneity and standardized to yield tissue-specific intensity values. All parameters for the fuzzy affinity relations were fixed for a specific input protocol. Scale-based fuzzy affinity was used to better capture fine structures. Brain tissues were segmented as 3D fuzzy-connected objects by using relative fuzzy connectedness. The user can specify seed points in about a minute and tracking the 3D fuzzy-connected objects takes about 20 seconds per object. All other computations were performed before any user interaction took place. Segmentation of brain tissues as 3D fuzzy-connected objects from MRI data is feasible at interactive speeds. Utilizing the robust fuzzy connectedness principles and fast algorithms, it is possible to interactively select fuzzy affinity, seed point, and threshold parameters and perform efficient, precise, and accurate segmentations.

Details of the new segmentation method are given in Section 4.2. Section 4.3 describes the methods of evaluation used in this work. Results and discussion are given in Section 4.4, and concluding remarks in Section 4.5. The method described in this chapter was published in [132].

4.1 Introduction to Image Segmentation Methods

The purpose of image segmentation were described and a classification of segmentation techniques were given in Section 3.1. Applications, wherein segmentation techniques using fuzzy connectedness principles [187] have been used successfully were also mentioned there. Among these large applications (involving several thousands studies) were several in which segmentation of brain images were required: MR brain image analysis for quantifying white matter lesions [113, 164, 189], brain tumor quantification [118].

However, there were several parts in these techniques where improvements can be achieved. All segmentation techniques depend on parameters. In previous applications of the fuzzy connectedness method, some of the parameters were set in advance, while others could only be extracted from the data being segmented, requiring human help. For example, parameters for the distribution of voxel intensities for a tissue were estimated from the small neighborhood of a few voxels that were specified by the user. This poses at least two problems. First, the reliability of the parameters is questionable, since it heavily depends on where the point was selected by the user. In case of fine,

thin structures, such as gray matter, selecting reliable regions may be difficult. Also, this way, several user sessions are required within the entire segmentation process: one for the parameter estimation, and another for verification. The latter may take place hours or even days after the first session.

Techniques have been developed to overcome these issues. A technique for standardizing MR image intensities has been devised [128, 135] (see Chapter 2). It allows parameter setting without the need for user input. Also, fast algorithms have been developed for computing fuzzy connectedness at interactive speeds [123] (see Chapter 3). If images having intensities with tissue specific meaning, such as standardized MR images, are utilized, most of the parameters for the segmentation method can be fixed once for all, all intermediate data can be computed before the user interaction is needed and the user can be provided with more information at the time of interaction. The reliable recognition (assisted by human operators) and the accurate, efficient, and sophisticated delineation (automatically performed by the computer) can be effectively incorporated into a single interactive process. This, we believe, is a major improvement in segmentation.

4.1.1 Segmentation Methods in MRI

Literature regarding image segmentation spans over several decades and a wide range of journals and collections. For reviews on image segmentation in general see [13, 139, 145], and for surveys and comparisons of methods for segmenting magnetic resonance images in particular see [22, 23, 24, 69, 90]. Since MRI is a very useful modality for studying brain tissue and function, a large body of MRI segmentation literature takes its application from brain studies. Images of a single protocol (commonly used T2/PD [73, 151], SPGR [65, 74, 152], or specialized MP-RAGE [82], AFFIRMATIVE [11]), or, often, a combination of images acquired with several protocols (T1/T2/PD [5, 75, 117], T2/PD/SPGR [7, 154], T1/T2/PD/FLAIR [40]) are utilized.

There are several ways to categorize (brain) (MRI) segmentation methods. We briefly give a few references to classes of different aspects (loosely following the categorization of [145]) and refer the reader to the review articles for further details and comparisons.

In most papers, the aim is binary (hard) segmentation wherein each voxel in the resulting image is determined to contain one particular tissue [75, 191]. On the other hand, in soft segmentation, voxels may be classified as containing a finite mixture of tissues and the percentage of each tissue in a voxel is given in the result as partial volume fractions [22, 69, 97, 106, 154, 175, 213] As opposed to the partial volume maps, the result of a soft segmentation may be some probability map of the tissue distribution [25].

According to their human interaction requirements, segmentation methods may be groupped as (fully) automatic [7, 37, 73, 153, 171, 175, 197] or interactive [50, 82, 166]. Interactive methods are rarely fully manual as they often utilize some level of automation and, thus, are referred to as semi-automatic methods.

Thresholding is one of the simplest operations on images and is almost never used on its own as a segmentation tool. However, it is often present in mixed algorithms, which also often utilize mathematical morphology [7, 95] and/or some variant of region growing [50, 65, 119, 154, 171] or clustering [203]. Classifiers and clustering algorithms partition a feature space derived from the input image data. Classifier methods require training data labeled by the user, therefore, they are supervised (semi-automatic) techniques [5, 11, 40, 75, 152, 191]. Clustering algorithms are unsupervised (automatic), since they do not use training data [22, 65, 69, 106, 117, 175, 191]. Statistical Markov random field (MRF) models may be utilized to support classification of voxels by modeling nonparametric distributions of tissue intensities, neighborhood correlations, and signal inhomogeneities [65, 73, 152, 197, 215]. Artificial neural networks (ANN) are also applied in classifiers wherein the weights of the ANN is determined by using the training data and the ANN is used to perform the segmentation [69, 103, 138, 153].

The watershed algorithm, as used in medical image segmentation, is a combination of edge detection and morphological operations that partition the image into homogeneous regions. It is

followed by a merging step wherein the small, oversegmented regions are combined into regions that correspond to the same anatomical structure [15, 166].

Deformable models are also actively researched and widely used in medical image analysis [107]. These techniques use parametric curves or surfaces to describe region boundaries. These curves (surfaces) deform under different constraints coming from the underlying image data (e.g., gradients) and from the model (e.g., shape, texture) [29, 31, 37, 85, 86, 108, 213].

In atlas-guided approaches, segmentation is achieved by registering a pre-segmented (labeled) atlas image to the image to be segmented and mapping the atlas labels to the registered image voxels [1, 41, 62]. The atlas is most often the Talairach brain atlas and usually non-rigid (also called elastic) transformation is found for registration.

4.1.2 Need for Image Segmentation Workshops

Despite the vast amount of literature and wide range of approaches to image segmentation, general, reusable frameworks seem to be lacking. General frameworks, that are readily applicable to images acquired with different protocols do not seem to have been published. Most papers deal with methods devised for specific tasks, tailored to a specific acquisition protocol and body region. In many cases, the segmentation method is a simple, *ad hoc* combination of basic tools or some more sophisticated combination involving heavy "engineering".

We emphasize the need for a *workshop*. That is, for a new protocol, the workshop allows quick experimentation to complete the set up phase quickly rather than requiring months and years of basic research and development. In this chapter, we describe such a workshop and its application to brain MRI segmentation. The method we devised using the general framework is an approach to the segmentation of the whole brain parenchyma (BP), as well as to the precise quantification of gray matter (GM) and white matter (WM) on several routinely acquired MR images of MS patients, such as fast spin-echo (FSE) proton density (PD) and T2-weighted, T1-weighted, and spoiled-gradient (SPGR) images. We do not attempt, however, to segment abnormal tissue regions, such as MS lesions, as separate objects in this chapter.

4.2 Interactive Segmentation Method

First, we give an overview of the segmentation method. It is followed by more detailed description of each step in subsequent sections.

Fuzzy connectedness constitutes a main component of the segmentation method described in this chapter. Principles of fuzzy connectedness were described in Section 3.2, and most of Chapter 3 dealt with efficient algorithms for computing fuzzy connected objects in images. In this chapter, we use those notions and assume the basic knowledge of the fuzzy connectedness framework.

4.2.1 Overview of the Segmentation Method

We distinguish two phases in this work. First, the segmentation procedure is set up, parameters are trained for, and fine tuned. This is performed only once for each protocol. Then, each dataset is segmented into the desired objects (i.e., BP, WM and GM). Below are the main steps of the segmentation method. We give details of each step in subsequent sections.

TS1-TS5: Training and set up
SS1: Correcting for RF field inhomogeneity
SS2: Standardizing MR image intensities
SS3: Computing fuzzy affinity
SS4: Specifying seed points and volume of interest (VOI)
SS5: Computing fuzzy connectedness

SS6: Creating the brain mask
SS7: Verifying the brain mask
SS8: Creating masks for each object
SS9: Computing volume

Most steps are automatic, their parameters are determined and fixed in the training/setup phase. For those steps, where user input is needed, we shall mention the amount of input and the way it is required. Depending on what acquisition protocol was used, small differences may be present as to how fuzzy affinity is computed. For example, when using dual-echo FSE images, both PD and T2 scenes, as well as a third, derived scene are utilized in the affinity computation, while for T1, for example, only one input scene is used. This difference may be considered hidden if we formulate it with vector valued scenes and affinities. However, in our implementation, affinities for each image are separately computed and then combined. See [217] for a fully vectorial formulation of fuzzy connectedness.

4.2.2 Training Phase

For the training phase, for each protocol, a few datasets are selected and used to extract the values for the parameters. These images are also used for testing the flow of operations and the control scripts step-by-step. The training procedure mostly requires continuous user control, since the fine tuning of parameters is a modify-and-verify iteration. The major steps of the training phase are the followings.

TS1: Perform the initial steps of the segmentation algorithm (SS1-SS2). SS4 (seed and VOI specification) may also be performed in advance. The reason for it being listed in the segmentation procedure after SS3 (affinity computation) is to make it possible to require only one user session. Otherwise, since affinity computation is a relatively slow operation, seed specification and brain mask verification would be possible only in two separate interactive sessions.

TS2: Determine threshold intervals to roughly segment the different tissues (CSF, WM, GM) in the intensity images. Threshold values for different input images should now be close to each other due to intensity standardization; thus, a common threshold can be determined. Compute the distribution of intensities within these segmented regions for each of the training data sets and compute a common value (average) for their parameters (mean and standard deviation). These values are good initial values for setting up the fuzzy affinity relations.

TS3: Perform the segmentation steps of the algorithm (SS3-SS8).

TS4: Evaluate the segmented objects. If necessary, modify the parameters and go to TS3. Otherwise, go to TS5.

TS5: Save the found best parameters for future use.

4.2.3 Correcting for RF Field Inhomogeneity

One of the well-known problems in MR imaging is intensity variation due to the RF field inhomogeneity. Most image processing and analysis methods are affected by this variation and their performance can be improved considerably if this variation is corrected. We applied a general, entirely image-based strategy for correcting this variation. The method is completely automatic, acquisition-protocol-independent, and requires no prior knowledge [216].

It is an iterative algorithm in which at each iteration a set of homogeneous tissue regions are obtained, no matter what each tissue region is. A smooth (2nd order polynomial) function is fit to the intensity variation within each tissue region and then combined into a single “gain field” for which the image is corrected. To determine homogeneous tissue regions, local scale values [163] are

computed in intensity standardized images [135]. Then, for the k largest mutually exclusive scale regions in the scale image, the mean μ_k, and standard deviation σ_k of the standardized intensities is determined within the corresponding regions of the standardized image. The standardized image is thresholded depending on these parameters to yield homogeneous regions. Standardization is effected at the end to make sure that the corrected images are not biased by the fitting strategy and that the intensity gray scale variations are also accounted for.

4.2.4 Standardizing MR Image Intensities

The image intensities in MRI do not have a fixed meaning, not even within the same protocol for the same body region obtained on the same scanner for the same patient. This poses problems in image display, segmentation, and quantification. The results of segmentation by two physicians are likely to differ because of the differences in their training. Setting values for the parameters for the non-manual methods becomes more difficult without the same protocol-specific intensity meaning.

The method described in Chapter 2 offers a simple way of transforming the images so that there is a significant gain in similarity of the resulting images. It is a two-step process consisting of a training step and a transformation step. This new transformation results in a standard scale for each protocol and body region. Standardized images can be displayed with fixed windows without the need of per case adjustment. The method facilitates improving the degree of automation of image segmentation. More importantly, extraction of quantitative information about healthy organs or about abnormalities can be considerably simplified.

4.2.5 Computing Fuzzy Affinity

For different acquisition protocols fuzzy affinity was computed slightly differently due to the differences of the input scenes (for example, there are single- and dual-echo protocols). However, the main scheme was the same for all protocols, scale-based affinity with Gaussian functions was used [163]. For computational simplicity, we used the six-adjacency relation for α, weighted by the appropriate voxel dimensions.

For the dual-echo FSE protocol, we consider the given vector-valued scene to be equivalent to two scenes $\mathcal{C}_{\mathrm{PD}} = (C, f_{\mathrm{PD}})$ and $\mathcal{C}_{\mathrm{T2}} = (C, f_{\mathrm{T2}})$. The *angle scene* $\mathcal{C}_{\mathrm{angle}} = (C, f_{\mathrm{angle}})$ is derived from $\mathcal{C}_{\mathrm{PD}}$ and $\mathcal{C}_{\mathrm{T2}}$ such that $f_{\mathrm{angle}}(c) = \tan^{-1} \frac{f_{\mathrm{T2}}(c)}{f_{\mathrm{PD}}(c)}$. This scene emphasizes CSF with bright intensities while suppressing other tissues [189]. For each of GM and WM, affinity was computed from both $\mathcal{C}_{\mathrm{PD}}$ and $\mathcal{C}_{\mathrm{T2}}$ scenes separately, and then combined into a single fuzzy affinity for the object using a simple $\min$ operator. For CSF, affinity was computed from $\mathcal{C}_{\mathrm{angle}}$.

For the T1 protocol, affinity was computed for GM and WM. However, a separate affinity for CSF was not used because CSF has similar intensities to that of bone, and peripheral CSF is indistinguishable from the enclosing bone, thus CSF was treated as background for the brain. For SPGR, affinity was computed for GM and WM, as well as for CSF.

4.2.6 Specifying Seed Points and Volume of Interest

At this point, user help is needed in the form of specifying seed points for each of WM, GM, and ventricular CSF. Four points for each of WM, and CSF, and 8 points for GM were input for each scene. During this same interactive session, the user is also asked to specify the first and last slices in which brain tissue is present. This volume of interest is utilized to exclude those parts of the image (mostly the bottom slices) that also contain regions which have similar affinity values to those of the objects of interest and may get connected to them during computation of fuzzy connectedness.

4.2.7 Computing Fuzzy Connectedness

Relative fuzzy connectedness [160] was used to determine the fuzzy objects for different tissues. A voxel c is assigned to that object with respect to whose seeds c's fuzzy connectedness value is the strongest.

4.2.8 Creating the Brain Mask

If we were using absolute fuzzy connectedness, the K-connectivity scenes would need to be thresholded to give binary masks for the objects. However, relative fuzzy connectedness directly outputs binary masks for the objects without requiring a threshold.

In this step, we combine GM, WM, and CSF masks to yield a brain mask (the intracranial volume). Then we apply some morphological operators. This is necessary for several reasons. In patient data, abnormal tissues may lead to very low (zero) affinity values within the normal tissue volume (e.g., WM lesions) and no object would claim those voxels in the "competition" of relative fuzzy connectedness. By morphological operation, these holes get filled and misclassification error reduced. Also, small, isolated clusters of voxels classified as one object get reassigned labels.

4.2.9 Verifying the Brain Mask

The "retouched" brain mask is presented to the operator for verification. The correction performed by the operator is mainly removing extracranial regions that may get connected to the brain via weak links of fuzzy connecedness.

The previous two steps (computing fuzzy connectedness and creating the brain mask) are so rapid (a few seconds per scene) that this and the previous interaction (seed and VOI specification) may be included in the same user session.

4.2.10 Creating Masks for Each Object and Computing Volume

GM, WM objects are recalculated within the corrected brain mask. Volume of each object is simply computed by multiplying the number of voxels within its binary mask by the voxel volume.

4.3 Evaluation

The presented image segmentation method was evaluated using images of MS patients. Image data is described first and methods of evaluation are given next.

4.3.1 Image Data

MRI head datasets of over 40 different MS patients and 10 normal controls acquired with several routinely applied MRI protocols were used in the experiments described below. The patients and datasets were randomly selected from our database that contains about 100 patients with about 4000 3D scenes comprising fast-spin-echo dual-echo (FSE) T2-weighted (T2) and proton-density (PD) weighted studies, T1-weighted studies without (T1) and with Gadolinium enhancement (T1E), magnetization transfer (MT) studies without and with saturation pulse, as well as spoiled-gradient (SPGR) studies.

The selected datasets were visually inspected and found to be of good quality having no severe artifacts. Although a few datasets contained visible movement artifact or MRI shading, we kept them for the experiment. Unfortunately, artifacts occur in real situations and we wanted to evaluate the segmentation method in non-idealistic, real situations. Most datasets were of MS patients, some

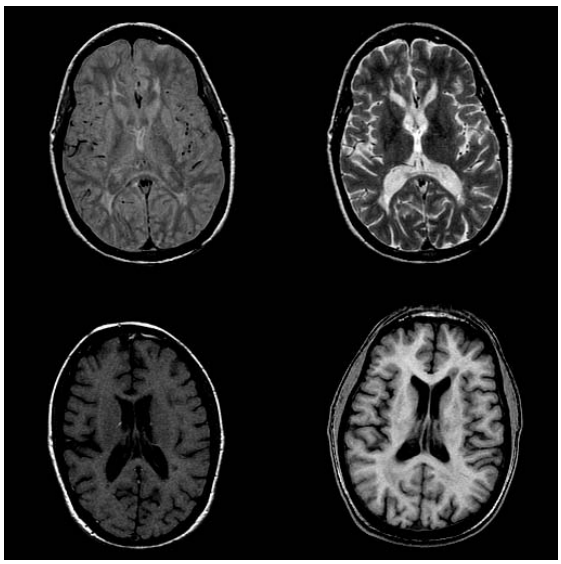


Figure 4.1: Slices of corrected and intensity standardized datasets. FSE PD, T2 (top), T1, SPGR (bottom).

of them with severe lesion load. We did not treat abnormal tissue (such as abnormal WM and MS lesions) as separate objects; these were segmented as part of WM or GM. Their separation can be accomplished in a subsequent step within each mask (e.g., separation of normal WM, abnormal WM, and lesions within the WM mask).

All datasets were acquired on the same GE Signa 1.5T scanner with a quadrature transmitter/receiver head coil and consisted of contiguous axial slices covering the entire brain with slice dimensions 256×256 and pixel size of $0.86\,\text{mm}^2$, and FOV = 22 cm. The FSE PD-T2 pairs were acquired with interleaved slices. All images were acquired at 3 mm slice thickness. The following protocol parameters were used: for FSE, TR = 2500 ms, with TE = 16–18 and 80–96 ms, NEX = 1; for T1, TR = 600 ms and TE = 27 ms, NEX = 1; for SPGR, TI = 500 ms and flip angle = 20, NEX = 2. For each patient, the 3 mm datasets consisted of about 50–60 slices. (See Fig. 4.1 for illustrative images.)

Repeated MR studies of several patients were selected for each of FSE, T1, and SPGR protocols. The two scans in the pairs for each patient were obtained with a short time gap in between during which the patients were removed from the table and then positioned again.

4.3.2 Precision (Reproducibility)

Precision may be measured with respect to several factors in the segmentation method. Both intra- and inter-operator precision, as well as repeat-scan precision were measured. For each test, volume difference and overlap agreement is measured. For repeat-scan overlap measurement, registration of the two scenes is necessary. Registration transformation was found between the original intensity scenes by using a method using mutual information [133]. The binary objects were transformed by this transformation.

4.3.3 Accuracy

Assessing accuracy of delineation is a difficult task due to the lack of a ground truth. Almost always some surrogates of truth, such as mathematical or physical phantoms, and manual tracing are used.

In all cases, the authenticity of the surrogate becomes questionable. Additionally, the precision with which the surrogate itself can be measured also becomes an important issue.

For the FSE datasets segmentation was readily available from previous research experiments. These segmentations were generated by a method and inspected by a neuroradiologist who manually corrected them where it was necessary to become the surrogate.

The segmentation results of our method were compared to the surrogate truth by using three accuracy measures [183]: *true positive volume fraction* (TPVF), *false positive volume fraction* (FPVF), and *false negative volume fraction* (FNVF).

4.3.4 Efficiency

We used an internal version of the 3DVIEWNIX [186] software system and PCs with Intel Pentium and AMD Athlon processors to do all image processing. Following the philosophy of 3DVIEWNIX, the software was written as a combination of several smaller programs, controlled by a script, instead of being a single, large, self-contained application. Due to this separate implementations, the programs communicate via files, that is, several file I/O operations are performed during the process, which adds to the processing time. If all steps are combined in a single application, these I/O operations are eliminated and the running time is reduced by several minutes in the automatic part, and by valuable seconds in the interactive step (on a modest performance PC).

Several types of efficiency measures may be considered. The running time (wall clock time) may be measured, although, it is highly dependent on what type of hardware the program is running on. Simply by switching CPU architecture (Intel, AMD, Sparc) or CPU clock frequency, wall clock running time may change substantially. We rather measure efficiency by the amount of the necessary human interaction during the process and also by the way this human input is required.

Our method has two phases. In the training phase, the user needs to interactively find the proper values of the parameters for the algorithms. This is a one time requirement for each protocol. In the real segmentation phase, user interaction is necessary when seed points and volume of interest are required and when the output of the algorithm needs to be verified and/or corrected. This is per-study interaction.

Seed points are specified by simple mouse clicks on the displayed images. The user may need to adjust the display window settings to better visualize the different tissues, although, when displaying intensity standardized images with pre-set windows, adjustments may not be necessary at all. The user slides through the axial slices for two purposes: to find good locations for the seed points, and, to specify the volume of interest, that is, to find the first and last slices containing brain tissue. During verification/correction, the user again slides through the axial slices of the original image. At this time, the corresponding slice of the binary brain mask is overlaid, and the user may cut misclassified regions from the mask or add missed regions to it, if found necessary.

4.4 Results and Discussion

After a few display examples of segmentation results, we illustrate by tables the performance of our new setup from several viewpoints: precision, accuracy, and efficiency.

4.4.1 Display Examples

Slices for original intensity scenes, fuzzy affinity and connectivity scenes, and binary segmentation results for different objects are displayed for each protocol in Figs. 4.2, 4.3 and 4.4.

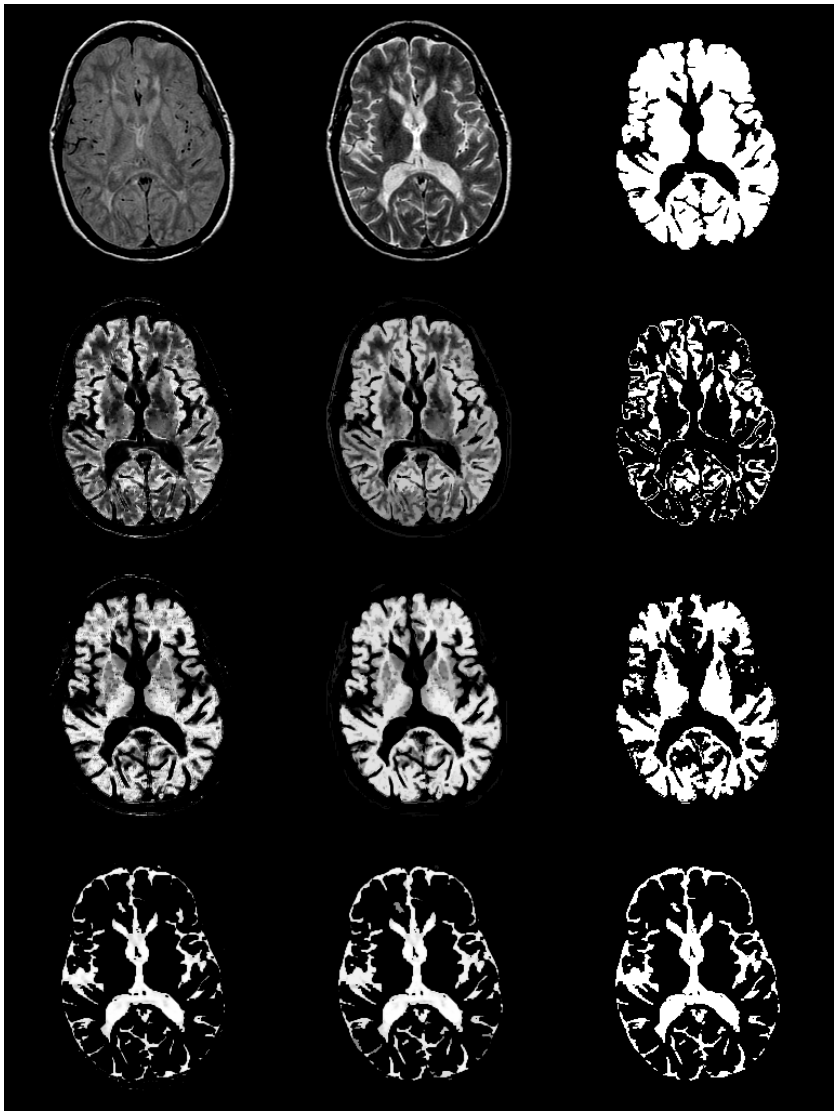


Figure 4.2: Slices from a standardized FSE PD, T2 study pair (top row, left and middle images), the corresponding slices from the scenes depicting the fuzzy affinity relations for the GM, WM, and CSF objects (first column), the same slices from the scenes depicting the connectedness values (second column), and the hard (binary) segmented objects (third column). Binary mask for brain parenchyma is shown in the top right image.

4.4.2 Precision (Reproducibility)

Tables 4.1 through 4.3 show repeat-scan, intra-, and inter-operator precision for segmenting brain parenchyma (BP), gray matter (GM), and white matter (WM) from images of different imaging protocols. Precision is measured between the segmentations from two scans by the same operator, two segmentations by the same operator, and segmentations by two operators. Volume difference and overlap agreements are expressed in percent. Mean and standard deviation over the image population is displayed.

Repeat-scan precision for the brain parenchyma is better than that for the separate objects. Vol-

Table 4.1: Repeat-scan precision of the method for segmenting BP, WM, GM, and CSF from images of different imaging protocols. Mean and standard deviation over the image population is displayed.

		FSE	T1	SPGR
volume difference (%)	BP	0.36±0.34	1.66±1.58	1.16±0.42
	GM	1.84±1.41	4.46±4.30	3.84±1.76
	WM	2.62±1.23	3.34±1.31	4.37±2.63
	CSF	2.45±1.68	5.72±3.58	3.90±4.47

Table 4.2: Intra-operator precision of the method for segmenting BP, WM, GM, and CSF from images of different imaging protocols. Mean and standard deviation over the image population is displayed.

		FSE	T1	SPGR
volume difference (%)	BP	0.02±0.02	0.09±0.06	0.11±0.06
	GM	0.03±0.02	0.14±0.08	0.19±0.14
	WM	0.03±0.03	0.04±0.04	0.03±0.03
	CSF	0.60±0.53	1.64±1.16	0.72±0.34
overlap agreement (%)	BP	99.96±0.03	99.91±0.05	99.92±0.04
	GM	99.96±0.03	99.86±0.09	99.86±0.08
	WM	99.96±0.04	99.97±0.02	99.98±0.02
	CSF	99.50±0.39	97.96±0.76	98.96±0.27

Table 4.3: Inter-operator precision of the method for segmenting BP, WM, GM, and CSF from images of different imaging protocols. Mean and standard deviation over the image population is displayed.

		FSE	T1	SPGR
volume difference (%)	BP	0.03±0.03	0.05±0.04	0.27±0.19
	GM	0.04±0.04	0.11±0.10	0.45±0.37
	WM	0.03±0.03	0.07±0.04	0.10±0.04
	CSF	0.47±0.37	1.94±1.73	1.88±1.04
overlap agreement (%)	BP	99.94±0.05	99.81±0.05	99.84±0.09
	GM	99.95±0.04	99.73±0.12	99.73±0.17
	WM	99.94±0.06	99.90±0.04	99.94±0.02
	CSF	99.28±0.41	97.32±1.13	98.48±0.24

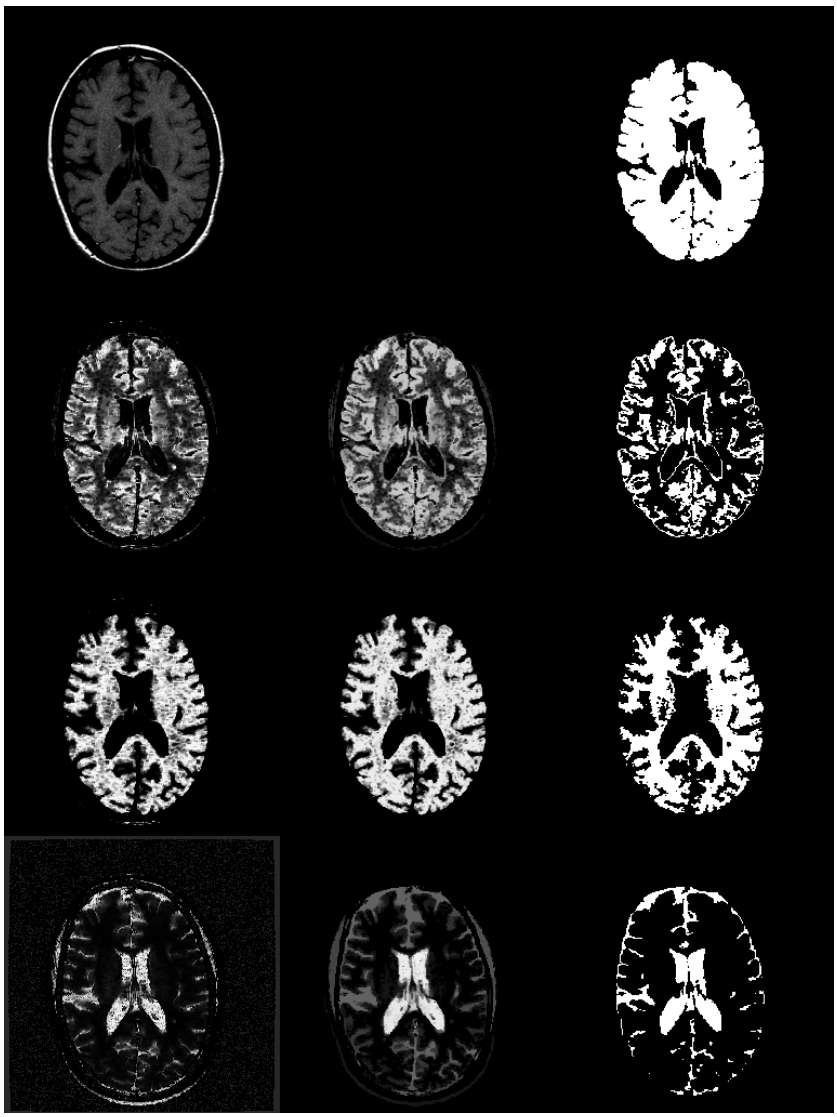


Figure 4.3: Slice from a standardized T1 study (top left image), the corresponding slices from the scenes depicting the fuzzy affinity relations for the GM, WM, and CSF objects (first column), the same slices from the scenes depicting the connectedness values (second column), and the hard (binary) segmented objects (third column). Binary mask for brain parenchyma is shown in the top right image.

ume disagreement is less than half percent for BP segmented from FSE and between 1 and 2% for T1 and SPGR protocols. As expected, the volume agreement is a few percent lower for GM, WM, and CSF than that for the whole brain. When segmenting FSE datasets, two scenes (proton density and T2-weighted) utilized in the segmentation, therefore, more (complementary) information is used that results in higher reproducibility. This is reflected in volume disagreements of the tissues.

Intra- and inter-observer precision is very high and comparable, 99% and above overlap and almost negligible volume difference between compared segmentations. The only exception is CSF for which volume disagreement is slightly higher and the overlap is "only" between 97.3% and 99.5%. This may be explained with the small CSF volume, compared to the other tissues, and also

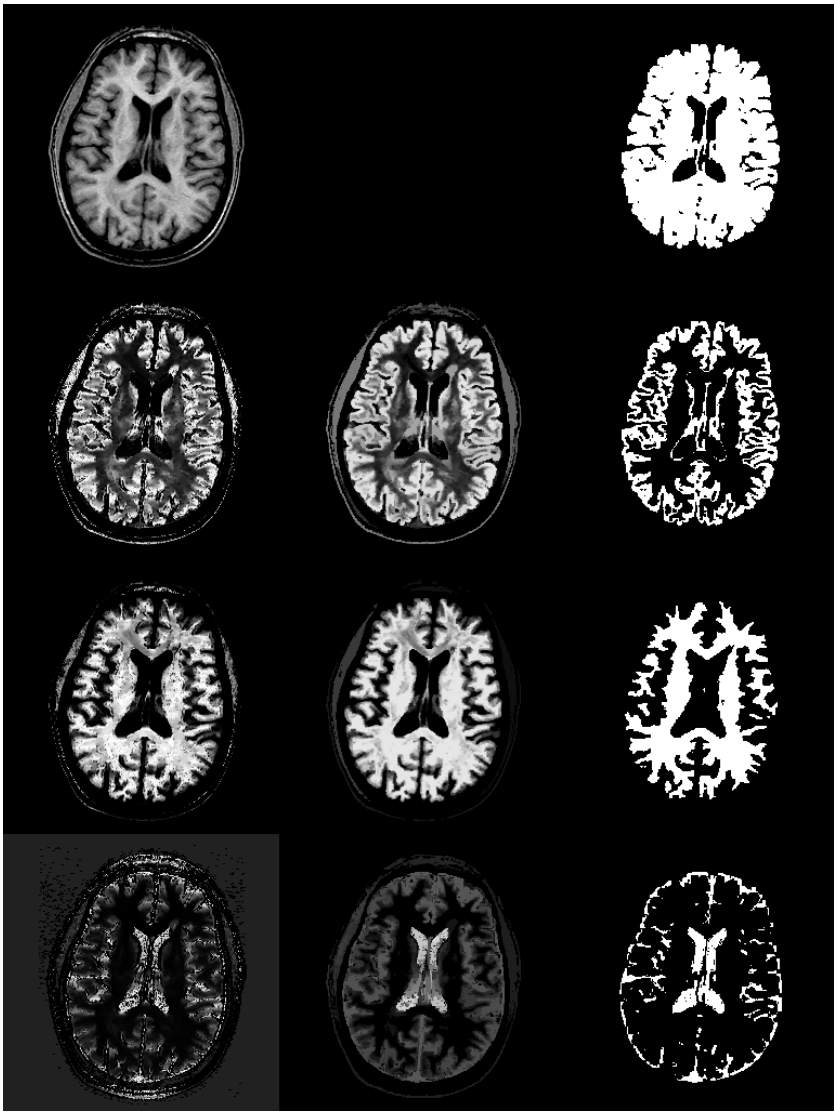


Figure 4.4: Slice from a standardized SPGR study (top left image), the corresponding slices from the scenes depicting the fuzzy affinity relations for the GM, WM, and CSF objects (first column), the same slices from the scenes depicting the connectedness values (second column), and the hard (binary) segmented objects (third column). Binary mask for brain parenchyma is shown in the top right image.

that the very thin peripheral CSF structures are much more than other tissues subjected to the brain mask correction that is the manual step of the method.

4.4.3 Accuracy

Table 4.4 shows FPVF, FNVF, and TPVF for segmenting brain parenchyma (BP), gray matter (GM), white matter (WM), and CSF from FSE images. TPVF, FPVF, and FNVF are with respect to the surrogate truth and expressed in percent. Mean and standard deviation over the image population is displayed.

Table 4.4: Accuracy of the method for segmenting BP, GM, WM, and CSF from FSE images. Mean and standard deviation over the image population is displayed.

	TPVF (%)	FPVF (%)	FNVF (%)
BP	99.82±0.28	1.31±0.67	0.18±0.28
GM	95.18±2.82	2.06±1.19	4.82±2.82
WM	98.91±1.03	6.61±3.79	1.09±1.03
CSF	93.27±4.14	1.41±1.28	6.73±4.14

The method is highly accurate with over 99% TPVF and less than 2% FPVF for the brain parenchyma, and slightly lower TPVF for the WM and GM objects. The slightly higher FPVF and FNVF for the WM and GM objects is due to misclassifications within the white matter near the ventricles. There is also some misclassification of GM/CSF partial volume voxels as WM around the brain parenchyma. Since this affects most slices and CSF volume is much less than that of BP, this results in a FNVF of around 7% for CSF.

4.4.4 Efficiency

The following time measurements were performed on an AMD Athlon XP processor running at 1400 MHz. These measurements are highly dependent on the CPU architecture and clock frequency used. The values given below are average values, individual studies may require a little more or less, depending on the noise, artifacts in the images, as well as the severity of abnormalities in the tissues. Running time (wall clock time) for step SS1 (inhomogeneity correction) is 2 minutes per scene for FSE PD, T2, and SPGR, and it is 3 minutes per scene for T1. Running time for steps SS2-SS3 for a FSE dataset is 3 minutes, that for a T1 or SPGR dataset is less than 2 minutes. Running time for step SS6 is a few seconds and it is included within the interactive session (SS5-SS7). Running time for steps SS8-SS9 is 5 seconds (again, dependent on computer power).

Next, we give time measurements needed for the interactive sessions. The one time training time for each protocol is a few hours. It is somewhat longer for the dual-echo FSE protocol than for the others. Specifying 4 seed points for each of CSF and WM, and 8 points for GM, and selecting the top and bottom slices of the volume of interest takes 90 seconds per study. Verifying and correcting the brain mask takes about 3 minutes per study. Most corrections take place at the bottom and the top of the brain.

4.5 Concluding Remarks

We presented a new method for segmenting brain MRI that is using the scale-based relative fuzzy connectedness framework. The method utilizes standardized MR images and fast algorithms, and thus, the user interaction may be reduced to a single interactive session. Once the appropriate parameters were fixed, the same procedure is applicable regardless of the MRI protocol. The segmentation method was developed within a general framework, a segmentation workshop. The workshop allows quick experimentation to complete the set up phase quickly rather than requiring months of basic research and development.

Segmentation of brain tissues as 3D fuzzy connected objects from MRI data is feasible at interactive speeds. Further, there is no need to devise protocol-specific segmentation strategies. Utilizing the robust fuzzy connectedness principles and fast algorithms, it is possible to interactively select

fuzzy affinity, seed point, threshold parameters and perform efficient, precise, and accurate segmentations. Also, utilizing the segmentation workshop, rather then starting from scratch, the setting of parameters for a particular protocol (or combination of protocols) may be done with a few hours or days of work.

Chapter 5

A Task-specific Comparison of 3D Image Registration Methods

We present a new class of approaches for rigid-body registration and their evaluation in studying Multiple Sclerosis via multi protocol MRI. Three pairs of rigid-body registration algorithms were implemented, using cross-correlation, normalized cross-correlation and mutual information, operating on original gray-level images and utilizing the intermediate images (called scale images) resulting from our new scale-based method. In the scale image, every voxel has the local "scale" value assigned to it, defined as the radius of the largest ball centered at the voxel with homogeneous intensities. 3D image data of the head were acquired from 10 MS patients using 6 MRI protocols. Images in some of the protocols were acquired in registration. The co-registered pairs were used as ground truth. Accuracy and consistency of the 6 registration methods were measured within and between protocols for known amounts of misregistrations. Our analysis indicates that there is no "best" method. For medium misregistration, the method using mutual information, for small and large misregistration the method using normalized cross-correlation performs best. For high resolution data the correlation method and for low resolution data the mutual information method, both using the original gray-level images, are the most consistent. The use of local scale information in fuzzy connectedness segmentation and image filtering has been previously demonstrated. Scale may also have potential for image registration as suggested by this work.

We describe the registration methods compared in this work in a general framework in Section 5.2. A brief description of the new scheme using object scale is given here as well as the details of our implementation. The clinical data and experiments are described in Section 5.3. Results and discussion are given in Section 5.4, and concluding remarks in Section 5.5. This chapter is based on [133, 134].

5.1 Introduction to Image Registration

The purpose of image registration is to represent information pertaining to the same object system of study in a common coordinate system. Registration becomes necessary primarily in the following situations:

(i) Scenes are acquired for the same body region from different modalities, for example, CT, MRI, fMRI and PET, for a patient's head. The need for registration comes from the fact that fMRI and PET provide mainly functional information and lack anatomic details, and vice versa for CT and MRI. Registration helps associating the functional activity information with specific anatomic regions.

(ii) When using different protocols within the same modality for the same body region, for example, T1-weighted images of the brain before and after administration of Gd-DTPA contrast

agent. Here registration helps to identify the regions that enhance and thereby to study certain disease processes.

(iii) Scenes are acquired for the same body region using the same modality for different time instances. The time instances may be close to each other for studying the motion or displacement of an object in the body region, or they may be far apart for studying longitudinally the growth or change in an object.

(iv) In certain interventional procedures, information derived from acquired scenes is utilized to provide navigational aids for the devices used in the procedure. In these situations, it becomes necessary to register the device, the body region, and the scene.

(v) Scenes acquired for a given body region are to be matched to a computerized atlas or model for the same body region. This is often helpful in studying statistically the variations in certain object measures over a subject population as well as in scene segmentation.

No registration is needed between two scenes if both the following conditions are satisfied:

(a) The same objects of study are observable in both scenes and they do not change shape.

(b) The relationship among them and to the scene coordinate system does not change.

For any objects, even static, in any body region, these conditions are difficult to meet, and therefore, registration becomes necessary in the situations described earlier.

Registration methods may be categorized in several ways. For excellent surveys covering and classifying most of the image registration literature, see [105, 196]. Broadly, registration methods may be divided into two groups: *scene-based* and *object-based*. In scene-based methods [211], the transformation needed to convert a given scene to match spatially with another given scene is estimated by matching the intensity pattern in the two scenes. In object-based methods [143], structure information derived from the two scenes is matched for this estimation. The registration transformation itself may be *rigid* [211] or *elastic* (also known as *deformable*) [8]. In the rigid case, the mismatch is assumed to be due to a global translation and rotation of one of the scenes/structures. Scaling, if required, is usually handled easily by an interpolation of one of the scenes prior to registration (or extraction of structures) to make its voxels equal in size to that of the other. In the elastic case, it is assumed that local deformations are needed in addition to a global translation and rotation, to achieve a match.

5.1.1 Registration as Applied to Studying MS

Much of the motivation for this investigation originated from our on-going research in MS images analysis. Therefore, we shall describe briefly in this section this basis as related to MS. Image analysis as applied to MS will be covered in more detail in Chapter 6.

The most commonly used metric to clinically measure disease progression in MS is the Expanded Disability Status Scale (EDSS) introduced by Kurtzke [92]. The clinical quantification of disease severity in this method is subjective and sometimes equivocal. The development of new treatments demands objective outcome measures for relatively short trials. Therefore, MR imaging has become one of the most important paraclinical tests for diagnosing MS and in monitoring disease progression in MS [100].

MR imaging has proven to be a very sensitive marker of the MS disease since macroscopic areas of damage or loss of myelin can be imaged with hyper or hypo intensity with respect to the surrounding tissues. A variety of protocols are being investigated for improving the conspicuity of the lesions in the images in the early inflammatory stage and in the advanced stage for both the microscopic and macroscopic processes. T1-weighted imaging with and without Gd-DTPA enhancement, dual spin-echo imaging providing the T2 and the proton density (PD) values for each image element, and magnetization transfer (MT) imaging are all utilized for characterizing and quantifying MS.

T1-weighted imaging may show lesions as hypo intense regions [190]. Post Gd-DTPA contrast T1-weighted imaging (T1E), however, shows hyper intense signals in the vicinity of the inflammatory regions [68]. In MS, Gadolinium-DTPA enhancement on MR images is characterized by blood

brain barrier (BBB) breakdown and intense inflammation which represents the acute stage of the evolution of MS [68, 71, 88, 120]. Enhancement is more sensitive than either clinical examination or T2WI in detecting disease activity and potentially can separate clinical groups [114, 169, 179]. T2 and PD images, on the other hand, both show the actual regions of the lesions with hyper intense values [165]. The sensitivity of T2 weighted imaging in the detection of MS has long been recognized [100, 158, 165]. The extent of lesions on T2 images, expressed usually by the total volume of lesions (T2LV), is currently regarded as the MRI measure of this disease burden. An important limitation of T2 imaging is its low pathological specificity [115]. All histopathological features of MS, such as inflammation, edema, demyelination, axonal loss, gliosis and remyelination are seen as hyperintense lesions while not all of these processes are clinically relevant.

Magnetization transfer (MT) images, especially the MT ratio (MTR) images, have been shown to provide information about the extent of both macroscopic and microscopic disease [36]. Therefore, MTR images offer the potential of quantifying the total extent of white matter loss that occurs in a diffused fashion. In MT imaging [210], two consecutive sets of images are acquired, one with off-resonance pre-saturation of the relatively immobile macromolecular protons and one without. A MTR image is then computed from the two images according to $(M_o - M_s)/M_o$, where M_s and M_o are pixel intensities with and without pre-saturation. Published studies [36, 42, 54, 99] suggest that the MTR values for white matter in normal subjects are highly reproducible within each institution with a variation of less than 2%. Several studies have demonstrated that the microscopic aspect of the disease can be characterized by the MTR values measured in different regions (white matter (WM), lesions (LS), the periphery of lesions) in a much more specific fashion than by T2LV [194].

Early identification of MS is crucial since effective treatment would be potentially most beneficial at the onset of the disease. However, the findings on the relationship between lesion load quantified from MRI and disease severity for established MS are less certain. (The term "lesion load" as used in the literature refers to the number of lesions and/or T2LV.) It is generally believed that this may be in part due to limitations in the methods utilized for quantitation, which in many studies are mostly manual or manually intensive.

It is clear that the various MR imaging methods provide somewhat disparate pieces of the information relating to the MS puzzle. The use of quantitative MRI in predicting the onset of MS and in assessing disease severity in established MS is actively pursued at present [48]. In all such efforts, image registration to combine the complementary information from the images acquired under different protocols becomes vital. Many of the image segmentation methods for MS require one or more preprocessing operations, including registration to handle multichannel features [64, 89, 117, 197, 202, 204]. Registration is also utilized in the evaluation of techniques that try to assign tissue-specific meaning to the numerical intensity values in commonly used protocols such as T1, T1E, T2 and PD [60]. New quantitative measures are also being developed from the images for each of several tissue regions including the whole brain parenchyma (BP), gray matter (GM), WM, LS and cerebrospinal fluid (CSF), for patients with relapsing remitting, secondary progressive, and chronic progressive MS. These methods also depend on image registration. In summary, the study of MS can greatly benefit from image registration. However, efforts in this direction in actual clinical studies are rather sparse and focussed mainly on observing enhancement in post-contrast images upon subtraction from registered pre-contrast images.

5.2 The Compared Registration Methods

In Section 3.2, we defined a scene $\mathcal{C} = (C, \mathbf{f})$ in a general, n-dimensional, vector-valued fashion. In this chapter, we assume that the scene domain C is three-dimensional and the intensity function f is scalar-valued.

5.2.1 General Scheme for Registration

In this study we confine ourselves to the evaluation of rigid-body scene-based registration methods. Here, we are looking for a coordinate transformation that maps coordinates of scene $\mathcal{C}_{i2}$ into coordinates in scene $\mathcal{C}_{i1}$ such that some similarity measure between $\mathcal{C}_{i1}$ and the transformed $\mathcal{C}_{i2}$ is maximal. Formally, in scene-based registration, the input is a pair of scenes $\mathcal{C}_{i1} = (C_{i1}, f_{i1})$ and $\mathcal{C}_{i2} = (C_{i2}, f_{i2})$ of the same body region and the output is a pair of scenes $\mathcal{C}_{o1} = (C_{o1}, f_{o1})$ and $\mathcal{C}_{o2} = (C_{o2}, f_{o2})$. Assuming that $\mathcal{C}_{i2}$ (the *test scene*) is to be matched with $\mathcal{C}_{i1}$ (the *reference scene*), the output scenes are such that $\mathcal{C}_{o1} = \mathcal{C}_{i1}$ and $\mathcal{C}_{o2} = T(\mathcal{C}_{i2}) = (C_{i1}, f_{o2})$, where T consists of a geometric transformation T_G to register $\mathcal{C}_{i2}$ with $\mathcal{C}_{i1}$ and an interpolation operation T_I to estimate scene intensities at the new locations for each voxel c in $\mathcal{C}_{i2}$. An appropriate scene-based or object-based interpolation method is utilized for this purpose. A majority of the existing scene-based registration methods are of the rigid type [104, 144, 195, 208, 211]. They can all be characterized more-or-less by the following process.

1. (Optionally) interpolate $\mathcal{C}_{i1}$ and $\mathcal{C}_{i2}$ to make voxels cubic and equal in size in both scenes.
2. Choose a *criterion function* G for determining the degree of mismatch between (the transformed) $\mathcal{C}_{i2}$ and $\mathcal{C}_{i1}$ and select a *stopping criterion* S.
3. Initialize T by setting it to an initial translation and rotation of $\mathcal{C}_{i2}$ to roughly align with $\mathcal{C}_{i1}$.
4. Optimize $G(T)$ as a function of T (and of $S(T)$) by incrementally modifying T and reevaluating $S(T)$.

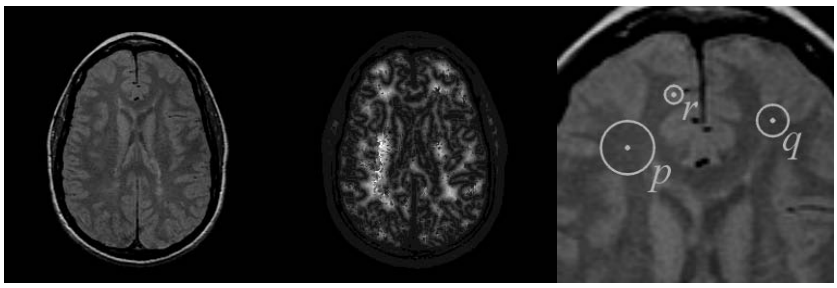
The essential differences in the different methods are in how the mismatch criterion function G and the stopping criterion S are selected and how $G(T)$ is optimized. Examples of criterion functions used include correlation coefficients of intensities or of intensity features [28, 104, 172, 173], variance of intensity ratios [172, 212], histogram dispersion [76], and mutual information [101, 206]. Search technique in Step 4 may use, for example, Powell's method [26, 96], the downhill simplex method [109], stochastic search [206], genetic methods [76], and (semi-)exhaustive search [140].

5.2.2 Scale as a New Feature

In scene-based methods, the mismatch criterion function G may be computed directly from the input scenes or from some feature scenes extracted from the original values. Several features (such as edges, ridges, troughs), have been proposed for scene-based image registration in the past [104]. We propose a new feature—the (local) object-scale—that has been recently used for aiding image segmentation [163] and filtering [161].

"Scale" is a fundamental, well-established concept in image processing [98]. Our idea of "scale", which is somewhat different from a similar concept used in computer vision [98], is to determine the size of local structures under a prespecified region-homogeneity criterion [163]. For example, in Fig. 5.1, local size of the structure to which voxel p belongs is bigger than that to which q or r belongs. There are other approaches to determine local scale measures. Pizer et al. [147] defined optimal scale at each spatial location as the scale for which "medialness" measure reaches a local maximum along the scale direction in the scale space representation. In their medialness model, scale represents the distance at which a pair of opposing boundaries are expected. In their interpretation, no continuity of homogeneity is used, whereas, in our approach, the concept is rather simple, object size is determined by looking for discontinuity of homogeneity in images, and the computation is simple [163].

We define *scale* in an image $\mathcal{C}$ at any voxel $c \in C$ as the radius $r(c)$ of the largest ball centered at c which lies entirely within the same object region (defined under a prespecified region-homogeneity



(a) (b) (c)

Figure 5.1: Illustration of local structure size (scale). An MRI slice of a MS patient's head (a), the corresponding scale image (b), and part of the original slice enlarged (c) to show details. In (b), scale values are not shown for the background part of the slice.

criterion). This adaptive scale concept has evolved in connection with fuzzy connectedness principles [189] and represents the local size of the object at the voxel. In [163], a simple and effective algorithm has been described which estimates $r(c)$ at every location $c \in C$ in any image $\mathcal{C}$ without explicit object segmentation but based on continuity of region homogeneity. It has been shown in [163] that this information of scale leads to a significantly improved object segmentation. For completeness, the algorithm for scale computation is reproduced below.

Let $\nu = (\nu_1, \nu_2, \nu_3)$ be the size of the voxels in the given scene $\mathcal{C} = (C, f)$, where ν_1, ν_2, ν_3 are the length of the voxel in the three principal directions in some measurement unit (mm in our application). A *ball* $B_{k,\nu}(c)$ of radius $k \geq 0$ and with center at $c = (c_1, c_2, c_3) \in C$ in $\mathcal{C}$ is defined by

$$B_{k,\nu}(c) = \left\{ e \in C \;\middle|\; \sqrt{\sum_{i=1}^{3} \frac{\nu_i^2 (c_i - e_i)^2}{\min_j [\nu_j^2]}} \leq k \right\}. \tag{5.1}$$

Note that $B_{k,\nu}(c)$ forms a digital ball when the spatial resolution is isotropic and a digital ellipsoid when it is not. The equation is formulated so that, in both cases, $B_{k,\nu}(c)$ represents a (quantized) ball in the scene domain. We will refer to $B_{k,\nu}(c)$ as a ball irrespective of whether or not ν is isotropic.

For a ball $B_{k,\nu}(c)$ of any radius $k > 0$ and centered at c, we define a fraction of object, $FO_{k,\nu}(c)$, that indicates the fraction of the ball boundary occupied by a region which is homogeneous with c, by

$$FO_{k,\nu}(c) = \frac{\sum_{e \in B_{k,\nu}(c) - B_{k-1,\nu}(c)} W_\psi(|f(c) - f(e)|)}{|B_{k,\nu}(c) - B_{k-1,\nu}(c)|}, \tag{5.2}$$

where $|B_{k,\nu}(c) - B_{k-1,\nu}(c)|$ denotes the number of voxels in $B_{k,\nu}(c) - B_{k-1,\nu}(c)$ and W_ψ is a homogeneity function. In the experiments described here, we use a zero-mean unnormalized Gaussian function for W_ψ. The algorithm for object scale estimation is presented in Algorithm 5.

The algorithm iteratively increases the ball radius k by 1, starting from 1, and checks for $FO_{k,\nu}(c)$, the fraction of the object containing c that is contained in the ball. The first time when this fraction falls below t_s, we consider that the ball contains an object region different from that to which c belongs. Following the recommendation in [163], we use $t_s = 0.85$.

To determine the standard deviation σ_ψ of W_ψ, we first compute the local in-plane intensity gradients (i.e., voxels c and d are in the same slice) over the entire scene domain. (We exclude local out-of-plane intensity gradients, since the out-of-plane resolution is usually coarser.) Here, local intensity differences are taken in both x and y direction over the entire scene and pooled together. Upper 10 percentile of the gradients are then discarded in order to account for inter object

Algorithm 5: Object scale extraction

Input: $\mathcal{C}$, $c \in C$, W_ψ, a fixed threshold t_s.
Output: $r(c)$.
begin
1 set $k = 1$;
2 **while** $FO_{k,\nu}(c) \geq t_s$ **do**
3 set k to $k + 1$;
endw
4 set $r(c)$ to $k - 1$;
5 output $r(c)$;
end

boundaries; the mean M_h and the standard deviation σ_h are then computed over the lower 90% gradients. Finally, σ_ψ is taken to be $M_h + 3\sigma_h$. In this fashion, the only parameter required in the method is determined automatically. A pair of neighboring voxels will be then assigned a value of intensity similarity depending on their intensity difference and the noise in the entire scene that is measured by σ_ψ.

Several hundred studies have been processed automatically and successfully using this method of scale estimation in our on-going projects for different protocol and body region combinations in several image segmentation applications. The scale image for the MR slice of an MS patient's head, shown in Fig. 5.1 (a), is shown in Fig. 5.1 (b). The image shows that interior regions have higher scale (brighter intensities) while detailed structured regions as well as boundary vicinities have lower scale (darker intensities). One set of our new methods used scale values directly as registration feature. In these cases, the registration method tries to match potential boundary regions to boundary regions, and homogeneous "flat" regions to flat regions. Incorporating boundary information in this manner into the registration method results in improved matching accuracy in certain cases. Our approach has the advantage of not requiring edge detection or explicit segmentation (which is a challenging task on its own). It uses information relating to the approximate whereabouts of the boundaries (regions of small scale) in a fuzzy manner.

5.2.3 Transformation

In our application (registering MRI head datasets for the purpose of computing MS disease-specific measures), the rigid-body assumption (i.e., the two scenes can be matched by global translations and rotations only) is satisfied well, since, in all cases, the images utilized in this study were acquired in the same session on the same scanner. Therefore, scanner calibration problems, geometric distortions as well as patient-related changes (due to disease evolution) can be neglected.

The geometric part T_G of a 3D rigid-body transformation T can be described by six parameters: three global translation components (t_x, t_y, t_z) and three rotation angles (ϕ_x, ϕ_y, ϕ_z) around the three coordinate axes. In our application, the origin of the coordinate system was placed at the center of the scene domain and the axes were defined by the natural directions of the digital scene. The transformation T_G is given by applying the three rotations and the translation component in the following order:

$$T_G = D \circ R_x \circ R_y \circ R_z \,, \tag{5.3}$$

where T_G represents the global (composite) transformation, D represents translation, and R_x, R_y, and R_z represent rotation around the x, y, and z axes, respectively, and $\circ$ denotes composition. For T_I, we used trilinear interpolation.

5.2.4 Matching Criteria

There are several measures proposed in the literature to define the degree of mismatch of two scenes. We utilized three different matching criteria to guide the search for optimum registration: cross-correlation, normalized cross-correlation and mutual information.

Correlation techniques [140, 195, 211] perform well in mono-modal registration wherein there is a linear relationship between the measurements for the same spatial elements in the two acquisitions. However, they are generally not the best candidates for matching images from different modalities because this special relationship is usually missing. In our application, however, the correlation method was expected to work well since, in all cases, we register MR images and we expect some relationship among the intensity values in the various MRI protocols utilized in our application.

When using object-scale feature for computing the mismatch function, that is, when registering the two scale images, the task can be considered as a mono-modal registration problem, since the scale values represent the same property. However, the different protocols show different contrast between tissues and the scale images may not show exactly the same objects always. For example, in PD images, GM and WM are portrayed well, and CSF is somewhat less prominent, while in T2 images, CSF is the most visible object, and the interface between WM and GM is usually blurred. As long as there are sufficient similar structures in the images, the matching algorithm performs well, even with some dissimilarities present in the images.

For any reference scene $\mathcal{C}_1$ and test scene $\mathcal{C}_2$, the correlation measure $CR(\mathcal{C}_1, \mathcal{C}_2)$ is defined as

$$CR(\mathcal{C}_1, \mathcal{C}_2) = \sum_{c \in C_1} f_1(c) \cdot f_2(c)\,, \tag{5.4}$$

where C_1 is the domain of the reference scene, and f_1 and f_2 are the intensity functions of $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. We confine the computation of $CR(\mathcal{C}_1,\mathcal{C}_2)$ to those voxels c which were within the head volume, thus excluding the large area of the more-or-less homogeneous background. The hypothesis for registration here is that the criterion function $G_{CR}(T) = CR(\mathcal{C}_1, T(\mathcal{C}_2))$ is maximized when the images $\mathcal{C}_1$ and $T(\mathcal{C}_2)$ are in spatial registration.

We define normalized correlation as

$$CR_n(\mathcal{C}_1, \mathcal{C}_2) = \frac{\sum_{c \in C_1} f_1(c) \cdot f_2(c)}{\sqrt{\sum_{c \in C_1} f_1^2(c) \cdot \sum_{c \in C_1} f_2^2(c)}}\,. \tag{5.5}$$

The criterion function for the normalized case is $G_{CR_n}(T) = CR_n(\mathcal{C}_1, T(\mathcal{C}_2))$.

Given that both images to be registered show information about the same underlying anatomy, there is mutual information between the two images. Mutual information (MI) is a basic concept from information theory, measuring the statistical dependence between two random variables or the amount of information that one variable contains about the other.

We treat the intensity values $a = f_1(c)$ and $b = f_2(c)$ in the two scenes that are to be registered as random variables A and B, respectively, over the scene domain. Estimations for the joint and marginal distributions of A and B can be obtained from the joint and marginal histograms of the two scenes. Mutual information is defined in terms of entropy in the following way:

$$MI(A, B) = h(A) + h(B) - h(A, B)\,, \tag{5.6}$$

where $h(A)$ is the entropy of the random variable A, and is defined as $h(A) = -\sum_a p(a) \ln p(a)$, $h(B)$ is defined analogously, and the joint entropy of two random variables A and B is $h(A, B) = -\sum_{a,b} p(a, b) \ln p(a, b)$. The criterion function in this case is simply defined as $G_{MI}(T) = MI(A, B')$, where A and B' are random variables representing intensities in $\mathcal{C}_1$ and $T(\mathcal{C}_2)$, respectively. Probability $p(a)$ for a given intensity value a in $\mathcal{C}$ in the above entropy formulation is estimated as the

number of voxels in $\mathcal{C}$ having value a divided by the size of C, that is,

$$p(a) = \frac{1}{|C|} \cdot |\{c \in C \mid f(c) = a\}| . \tag{5.7}$$

MI as a mismatch measure for image registration has been investigated by several groups [101, 206, 208]. The published results show that MI gives one of the best registration results in multi-modal image matching. This technique does not require any *a priori* model of the relationship between the modalities. It assumes only that one volume provides the most information about the other (i.e., $G_{MI}(T)$ is maximized) when they are correctly registered.

Having $G_{CR}(T)$, $G_{CR_n}(T)$, $G_{MI}(T)$, and letting $\mathcal{C}_1$ and $\mathcal{C}_2$ be also scale images of the given reference and test scenes, we consider the following six strategies for registration:

1. correlation of original scenes (c,o),
2. correlation of scale scenes (c,s),
3. normalized correlation of original scenes (cn,o),
4. normalized correlation of scale scenes (cn,s),
5. mutual information in original scenes (m,o),
6. mutual information in scale scenes (m,s).

5.2.5 Initialization

In almost all practical cases, the functions we try to optimize have several local optima. Assuming that reaching the global optimum is desirable, good initialization close to the global optimum is therefore very crucial for an iterative optimization process to avoid local optima as much as possible. In most cases, the two scenes to be registered cover almost the same body region and have similar orientation by acquisition. However, for different acquisitions, the field of view may be different and repositioning errors always occur. Usually, one of the images can be automatically corrected for scale differences by interpolation. Identifying the same position (volume of interest) and orientation, however, usually requires manual help from the operator.

We initialized the transformation T by aligning the centroids and the principal axes of the two scenes. Sometimes this principal axes method itself provides the final registration [180], but often it is only the initialization step to a more sophisticated registration method. The background (outside the head) was segmented by thresholding, and the foreground was used to compute the centroid and the principal axes using the following method.

Let ST_1 and ST_2 denote the two structures (objects) to match. Let $\{Q_1^i, \ldots, Q_{n_i}^i\}$ be the set of points representing ST_i for $i = 1, 2$, where each Q_j^i is a 3-component vector, representing the coordinates of Q_j^i. Then the centroid O_i of ST_i is given by

$$O_i = \frac{1}{n_i} \sum_j Q_j^i . \tag{5.8}$$

The 3×3 covariance (or inertia) matrix M_i of ST_i is given by

$$M_i = [Q_1^i - O_i, \ldots, Q_{n_i}^i - O_i][Q_1^i - O_i, \ldots, Q_{n_i}^i - O_i]^T . \tag{5.9}$$

The eigenvectors of M_i define the directions of the principal axes of ST_i. The axes system defined by O_i and these eigenvectors is called the *principal axes system* of ST_i. If ST_1 and ST_2 are similar in shape, then to match ST_2 to ST_1, O_2 is translated to O_1 and the principal axes of ST_2 are rotated to match with those of ST_1. Note that if x is an eigenvector of M_i, then so is $-x$. Hence, the positive and negative directions of the principal axes cannot be ascertained. If the rotations are known to be less than 90°, this knowledge can be utilized in estimating correctly the rotations. Otherwise, additional information becomes necessary to select one among the several solutions provided by

this method. Since we knew that the orientations of the original scenes are similar, the eigenvectors were selected such that the least amount of rotation is involved in the initial match.

There may be other problems with this initialization. For example, when the eigenvalues are very similar, it is possible that the wrong axes are matched and large rotations are introduced. We prevent this by a constraint on the rotation angles, and if the program detected such a problem, it swapped the principal axes.

5.2.6 Search Strategy

One of our previous implementations for multimodal image registration used exhaustive search to minimize the cross-correlation function [140]. Translations were varied in one voxel steps and rotations in a quarter degree steps. Since this approach involves evaluation of the correlation function for a large number of positions, it was computationally expensive. To reduce computation but still have the necessary capture range and precision, we used a multi-level approach. Multi-level methods are widely used in the literature for registration [102, 140, 174]. They have proved effective in reducing computation time as well as in helping to avoid local minima. In this study, we used a three-level Gaussian pyramid, each level being twice as coarse in resolution as the previous, level 0 being the original resolution. However, since our data are highly anisotropic (in z direction), we applied filtered subsampling within slices only. Thus, at the coarsest level (level 2), the data resolution was approximately isotropic.

At each hierarchy level, an optimal solution was determined which was used as the starting point for the next level. At the lowest resolution, search was initialized by the principal axes method. Powell's multidimensional direction set method [149] was used to optimize the mismatch function by using Brent's one-dimensional optimization algorithm for line minimizations. At each hierarchy level, the direction matrix is initialized with translation steps corresponding to the in-plane size of one voxel at that resolution, and the rotation step size is chosen such that it corresponds to one translation step at 60 mm from the center of rotation. Following the suggestions of [101], we used the parameter ordering $(t_x, t_y, \phi_z, \phi_x, \phi_y, t_z)$, so that images get aligned within the plane first, which may facilitate the optimization of the out-of-plane parameters. The fractional tolerance for Brent's line minimization method was set to $2 \cdot 10^{-4}$, that for Powell's method to 10^{-3}, and the maximum number of iterations in one step was 25 and 10 for Brent's and Powell's method. The translation components were expressed in terms of ν_1 ($= \nu_2 = 0.86$ mm in all our data) and the rotations in degrees. Double precision floating point arithmetic was used throughout the optimization process.

5.3 Evaluation

The presented image registration methods were compared using images of MS patients. Image data is described first and methods of comparison are given next.

5.3.1 Image Data

MRI head datasets of 10 MS patients (2 male and 8 female, age between 36 and 50 years, mean 43, SD 4.5) were used in the experiments described below. The patients and datasets were randomly selected from our database which contains about 100 patients with about 4000 3D scenes comprising fast-spin-echo dual-echo (FSE) T2 and PD studies, T1 and T1E studies, and MT studies. For each patient, six datasets comprising FSE T2 and PD, T1 and T1E, and MT without and with saturation pulse (referred to as MT1 and MT2 studies below) were selected. The datasets of each patient were acquired in a single session. FSE PD and T2 scenes are in registration by acquisition and so are the two MT scenes. The selected datasets were visually inspected and found to be of good quality having no severe artifacts. Although one dataset contained visible movement artifact, we kept it

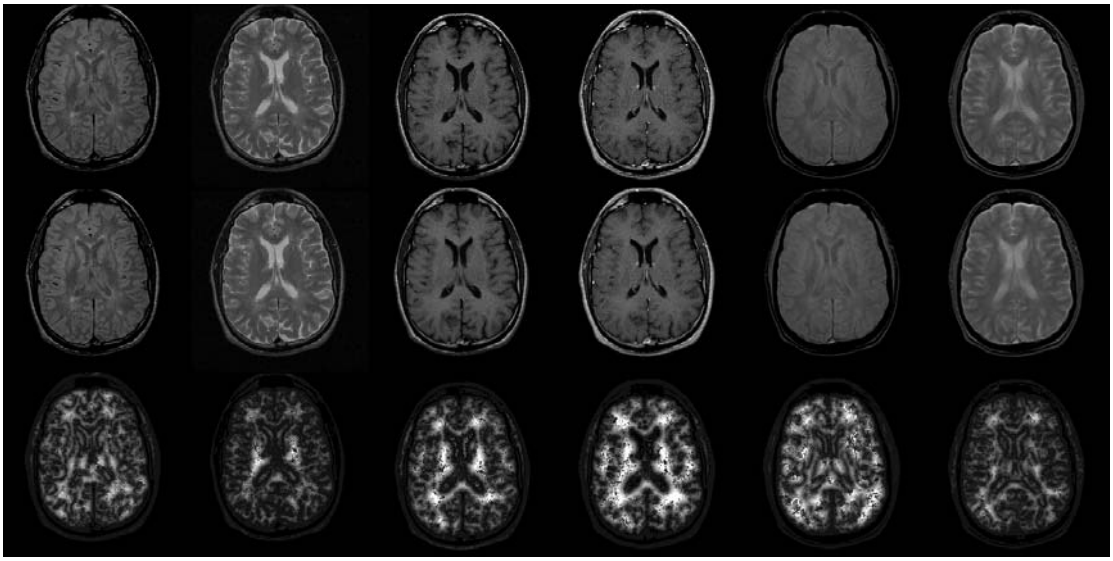


Figure 5.2: Original corresponding slices of each of the six datasets (FSE PD, FSE T2, T1, T1E, MT1, MT2) of one patient before registration (first row) and after registration (second row). Slices of the scale images corresponding to those in the first row are shown in the third row. (Scale values are not shown for the background part of the slice.) Note the similarity of the scale images across all protocols.

for the experiment. Unfortunately, artifacts occur in real situations and we wanted to evaluate the registration methods in non-idealistic real situations.

All datasets were acquired on the same GE Signa 1.5T scanner and consisted of contiguous axial slices covering the entire brain with slice dimensions 256×256 and pixel size of 0.86 mm^2, NEX = 1 and FOV = 22 cm. The MT images were acquired at 5 mm slice thickness while the other four were acquired at 3 mm. For all patients, for the MT image acquisition TR = 106 ms and TE = 5 ms, for the T1-weighted studies (with and without enhancement) TR = 600 ms and TE = 27 ms. For the FSE acquisitions, for all patients TR = 2500 ms, with TE = 16–18 and 80–96 ms. For each patient, the four 3 mm datasets consisted of the same number of slices (45–58), all 5 mm datasets included 28 slices. Slices of images of the six protocols from one patient are displayed in Fig. 5.2.

5.3.2 Experiments

The scenes were misregistered by known amounts of translation and rotation (3-3 each). The pairs that were in registration at acquisition were used as ground truth, and the accuracy and consistency of the 6 registration methods were measured within and between protocols for small, medium and large misregistrations, thus resulting in a data set of 540 3D scenes (9×6×10). (The extents of these misregistrations are described later.) Slices of these misregistered images for one dataset are shown in Fig. 5.3 for all nine cases.

To assess *intra-modality accuracy*, we used the following procedure. We take a scene $\mathcal{C}$ from each protocol in the set {PD, T2, T1, T1E, MT1, MT2}, and reslice $\mathcal{C}$ to create an artificially misregistered version, $\mathcal{C}_k = T_k(\mathcal{C})$, by applying a known transformation T_k. For each method $m \in$ {(c,o), (cn,o), (m,o), (c,s), (cn,s), (m,s)}, we find transformation T_{k_m} which registers $\mathcal{C}_k$ with $\mathcal{C}$, and compare T_k and T_{k_m}.

Using the pairs (PD, T2) and (MT1, MT2) which were in registration at acquisition, we can assess *inter-modality accuracy* in a similar way. The only difference is that, the resliced scene is registered

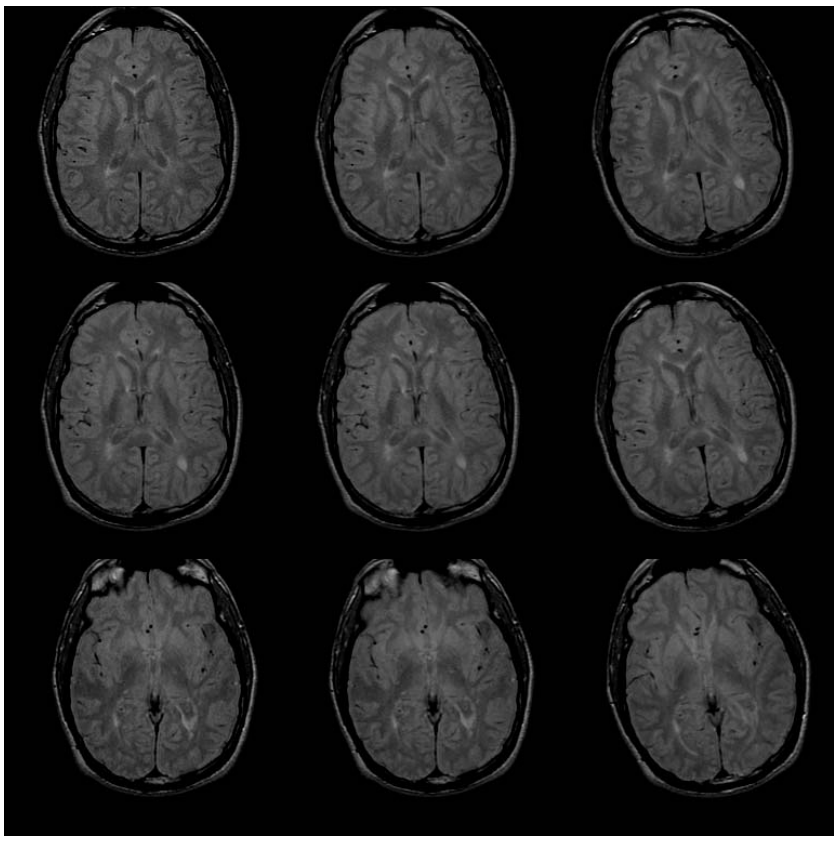


Figure 5.3: Slices of original and artificially misregistered data sets (PD scenes). Zero, small, and large rotations across columns, zero, small, and large translations down across rows.

to the other scene from the pair (e.g., rotate and reslice the T2 scene and then register to the PD scene).

Inter-modality consistency can also be assessed by utilizing the registered pairs. Register a scene $\mathcal{C}$ of protocol $\mathcal{P}$ to both scenes from a pair (e.g., register a T1 scene to a T2 and to the corresponding PD scene) and compare the two resulting transformations. This can be repeated for $\mathcal{P} \in \{$T1, T1E, MT1, MT2$\}$ by using the PD-T2 pair as reference, and for $\mathcal{P} \in \{$PD, T2, T1, T1E$\}$ by using the MT1-MT2 pair as reference. From the three groups together, we have a total of 54 experiments, each involving 10 data sets.

5.3.3 Figures of Merit

In each experiment, we compare two transformations. In the *accuracy* tests, these are the known misregistration and the transformation found by the method in case, and they are expected to be the inverse of each other. In the *consistency* tests, the two registrations to scenes of the registered pair are compared and they are expected to be equal. We measure the error between two transformations by the mean of the root mean squared (RMS) error for the 8 corner voxels of the box approximately bounding the head, i.e., the volume of interest in our application.

We combined three different rotations (zero, medium, and large angle) and three translations (zero, medium, and large displacement) to introduce 9 different misregistrations. All non-zero translations and rotations were such that all three directions/axes were involved: rotations were

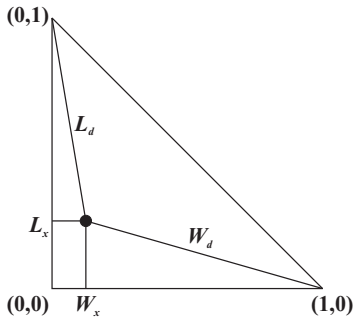


Figure 5.4: Mapping the number of wins, losses, and nonsignificant differences into a single "goodness" value for comparison. See text for details.

about 2.5 and 10 degrees, respectively, and translations were about 5 pixels (4.3 mm) and 25 pixels (21.5 mm), respectively. To reduce the number of comparisons and possibly improve statistics, we divide these nine transforms into three groups of misregistrations. After examining the results for each transform, we decided to combine those involving the same amount of rotation, since the methods performed with similar errors within these groups.

We devised a ranking system for the methods as follows. Each registration task is performed on datasets of 10 patients. For each method and misregistration (and patient data), for each comparison (i.e., accuracy or consistency test, see above), we computed the mean of the RMS error for the 8 corner voxels of the box approximately bounding the head. We can use the paired t-test to compare the errors of two methods performing the same task (same misregistration, same protocol, 10 different datasets). Based on the mean RMS error, one of the two methods is either significantly better than the other (i.e., smaller error), or there is no significant difference between the two in performing that particular task. For each misregistration group, for each reasonable pair of methods, for each task (protocol, accuracy/consistency test), we count the number of occurrences of significant wins, losses, and nonsignificant differences of the methods in question. Based on a combination of these counts, we rank the methods within a particular task. Let us consider a particular triplet of wins, losses, and nonsignificant differences (w, l, n). The total number of comparisons for this particular case is the sum $X = w + l + n$. We normalize the number of wins and losses to get values in the range $[0, 1]$: $W_x = w/X$, $L_x = l/X$. A particular configuration of wins and losses can be identified by a point with coordinates W_x and L_x in the normalized win-loss plane (Fig. 5.4). The closer this point is to the $(1, 0)$ point (all wins) and the farther it is from the $(0, 1)$ point (all losses), the better the configuration is. We use the following formula to compute the "goodness" value (γ) of a particular win-loss configuration:

$$\gamma = \frac{L_d}{W_d} = \frac{\sqrt{(1 - L_x)^2 + W_x^2}}{\sqrt{(1 - W_x)^2 + L_x^2}}. \tag{5.10}$$

5.4 Results and Discussion

In the majority of the 5400 registrations performed, RMS error of the resulting transformation was within the subvoxel range. In a few cases, RMS error was larger than the voxel size and there were a few situations in which the method did not find a reasonable registration at all. When applying any method to a large number of real images there is a chance that failure happens. The success of a registration method cannot be guaranteed, and in case it fails special care is required. Most of these failure cases corresponded to registering scale images with large misregistration (rotation).

Table 5.1: Comparison of methods for intra-protocol registration. The goodness values γ are listed.

thickness	method	small	medium	large	total
3 mm data	(c,o)	2.188	0.192	0.046	0.344
	(cn,o)	1.991	0.117	21.625	1.550
	(m,o)	3.222	421.000	6.250	10.346
	(c,s)	0.353	1.230	0.457	0.580
	(cn,s)	0.192	1.230	0.675	0.568
	(m,s)	0.707	1.000	0.812	0.831
5 mm data	(c,o)	1.000	0.310	0.053	0.322
	(cn,o)	4.281	0.831	inf*	5.200
	(m,o)	1.625	2.500	2.643	2.188
	(c,s)	0.233	0.737	0.068	0.279
	(cn,s)	0.737	1.000	0.500	0.744
	(m,s)	0.850	1.991	2.000	1.457

* The method significantly won in all cases, therefore, the denominator in the formula for γ is 0, thus, γ is infinity.

Table 5.2: Comparison of methods for inter-protocol registration. The goodness values γ are listed.

thickness	method	small	medium	large	total
3 mm data	(c,o)	1.355	1.176	1.000	1.168
	(cn,o)	0.400	0.615	4.281	0.940
	(m,o)	1.153	0.737	0.867	0.907
	(c,s)	1.413	1.153	0.737	1.053
	(cn,s)	1.000	1.153	0.615	0.907
	(m,s)	1.000	1.355	0.867	1.050
5 mm data	(c,o)	1.355	1.355	0.831	1.176
	(cn,o)	1.000	1.991	1.737	1.510
	(m,o)	1.625	1.991	2.500	1.991
	(c,s)	0.575	1.355	0.737	0.850
	(cn,s)	0.502	0.076	0.502	0.310
	(m,s)	1.355	0.707	0.867	0.949

For the 3 mm datasets, RMS error was around 0.5 mm for misregistrations with rotation and around 1 mm for the zero rotation cases, most of the error coming from translation in the (low resolution) z direction. RMS error was slightly higher (around 1 mm and 3 mm, respectively) for the 5 mm datasets.

Our analysis is based on all possible sensible comparisons (135 in all). The "goodness" values γ are shown in Tables 5.1 (intra-modality), 5.2 (inter-modality), and 5.3 (consistency). Values for the 3 mm datasets are shown in the top half of each table while those for the 5 mm datasets are in the bottom. Rows correspond to the methods, columns correspond to the misregistration groups (small, medium, and large), and the combined values over all groups (total).

Our overall observation is that there is no "best" method that stands out in all cases. In total, there is significant difference in performance for intra-protocol registration while for inter-protocol registration the best methods have more similar performance. The most competitive among the

Table 5.3: Comparison of methods for consistency. The goodness values γ are listed.

thickness	method	small	medium	large	total
3 mm data	(c,o)	2.828	1.153	1.286	1.555
	(cn,o)	0.502	0.928	1.991	0.972
	(m,o)	1.071	0.933	0.737	0.909
	(c,s)	1.679	1.071	0.737	1.080
	(cn,s)	0.557	0.801	0.801	0.717
	(m,s)	0.801	1.153	1.000	0.976
5 mm data	(c,o)	1.679	2.828	1.625	1.950
	(cn,o)	1.176	2.479	1.383	1.548
	(m,o)	2.828	2.828	1.355	2.142
	(c,s)	0.737	0.502	0.801	0.675
	(cn,s)	0.270	0.169	0.502	0.296
	(m,s)	0.850	0.535	0.801	0.722

methods seem to be (cn,o) and (m,o). We note that there are situations wherein (cn,o) is better than (m,o) and vice versa, and (m,s) is better than (m,s) is better than (m,o) and vice versa.

Method (m,o) using mutual information and the original intensities performs best on intra-modality registration for cases with medium misregistration, and for high-resolution data with small, and low-resolution data with large misregistration. However, method (cn,o) using normalized cross-correlation and the original intensities is better for cases with large misregistration and for low-resolution data with small misregistration. In total, method (m,o) performs better for the 3 mm data and method (cn,o) for the 5 mm data. Methods using scale images did not get a high score; however, in the large misregistration groups, method (m,s) ranks third. Similarly, in inter-modality registration, method (m,o) and (cn,o) beat the others in most groups, however, the scores are not that polarized. The mutual information method on scale images similarly ranks among the top three in four groups.

As far as consistency is concerned, the correlation method using original intensities ranks first in all groups for the 3 mm data and the mutual information method is more consistent for the low resolution data. Methods using scale images seem to be considerably less consistent than the others on low resolution data, according to our measure.

Our current implementation requires that both reference and test scenes have identical level of discretization and size. It was not affecting the results, since the original datasets were "compatible" and there was no need for extra interpolation. Another implementation limitation is that multiresolution pyramid is computed only in x-y directions, such that at the lowest resolution level, the reformatted data are close to isotropy. We chose this implementation to compensate for the highly anisotropic data.

5.5 Concluding Remarks

Three pairs of rigid-body registration algorithms were implemented, using cross-correlation, normalized cross-correlation and mutual information, operating on original gray-level images and on the intermediate images (called scale images) resulting from our new scale-based method. The scale images, are derived directly from the original images.

3D data of the head were acquired from 10 MS patients using 6 MRI protocols (T1, T2, PD, T1 with Gd, MT on, MT off) and both intra- and inter-protocol registration results were analyzed

among the methods. Our analysis indicates that there is no "best" method, although there are significant differences in the performance of the methods, for the particular application of combining information from multiprotocol MR imagery for studying MS.

Our results showed that object scale information (like a set of intrinsic markers) can be used in image registration to provide good results. The use of local scale information in fuzzy connectedness segmentation and image filtering have been previously demonstrated [161, 163]. Scale may also have potential for image registration as suggested by this work.

Chapter 6

MR Image Analysis in Multiple Sclerosis

Although imaging modalities such as computerized tomography have been studied for their effectiveness in capturing information about MS, the most ubiquitous modality at present is magnetic resonance (MR) imaging (MRI). The aim of this chapter is to give an overview of the techniques developed in recent years for quantifying MS via MRI. As described later in this chapter, a variety of MRI protocols have been utilized in understanding this disease (see Fig. 6.1). A brief background on MS is given in Section 6.1. We outline a general framework for MR image analysis in Section 6.2 and then examine the different methods developed for each MRI protocol in Sections 6.3, 6.4, 6.5, and 6.6. Throughout this chapter, phrases such as "we", "us", "our" refer to the Medical Imaging Section, Department of Radiology, University of Pennsylvania.

In an attempt to eventually replace EDSS (a widely used but rather subjective measure) by an objective measure to assess the natural course of the disease and its response to therapy, we have developed image segmentation methods based on fuzzy connectedness to quantify various objects in multiprotocol MRI. These include the macroscopic objects such as lesions, the grey matter (GM), white matter (WM), cerebrospinal fluid (CSF), and brain parenchyma as well as the microscopic aspects of the diseased WM. Fuzzy connectedness techniques for different protocols are described in Section 6.7 and their evaluation in Section 6.8. In Section 6.9, we summarize the lessons learned from this experience and the outstanding problems that need to be addressed. This chapter is based on [129, 185].

6.1 Introduction to the Study of Multiple Sclerosis

Multiple Sclerosis (MS) is an acquired disease of the central nervous system (CNS) characterized primarily by multifocal inflammation and destruction of myelin. Inflammation and edema are accompanied by different degrees of demyelination and destruction of oligodentrocytes, and may be followed by remyelination, axonal loss and/or gliosis. The disease was first characterized by Charcot [21] and has since been investigated extensively. The highest frequency of MS occurs in northern and central Europe, Canada, the USA, and New Zealand and South of Australia [91]. In the US, it affects approximately 350,000 adults and stands out as the most frequent cause of non-traumatic neurologic disability in young and middle-aged adults [70]. In its advanced state, the disease may severely impair the ambulatory ability and may even cause death.

MS is usually classified into three subtypes: (1) Relapsing-remitting (RR): Clearly defined disease relapses with full recovery, or with sequelae and residual deficit upon recovery, but there is no disease progression between relapses. (2) Secondary-progressive (SP): Initial RR disease course

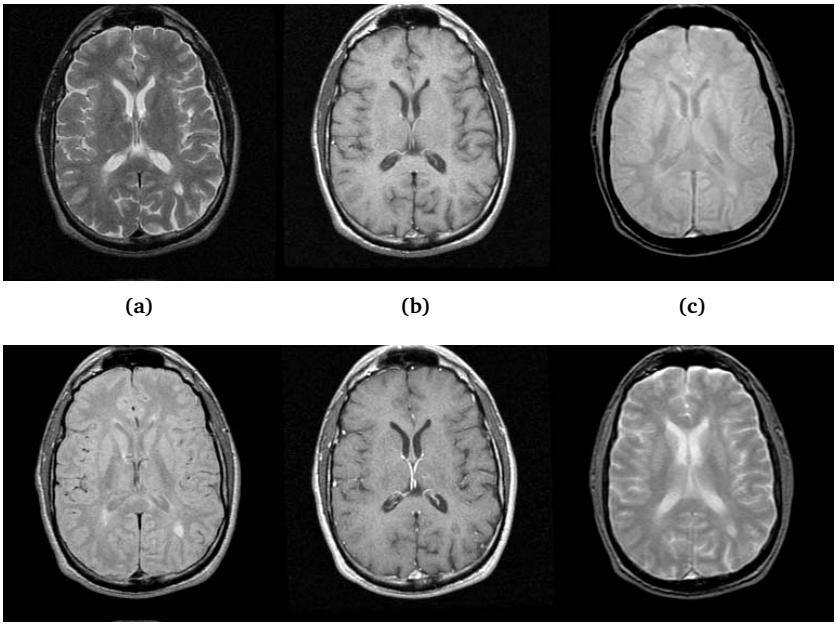


Figure 6.1: Images of an MS patient acquired with six protocols T2 (a), PD (d), T1 (b), T1 with Gadolinium enhancement (e), magnetization transfer with saturation pulse off (c) and on (f). The data sets were registered in 3D to show the same slice plane for all protocols.

followed by gradual progression with or without occasional super-imposed relapses. (3) Primary-progressive (PP): Gradual progression from the onset with occasional plateaus and temporary minor improvements.

There are several clinical measures commonly used to express the severity of the disease. These measures are subjective and may have reproducibility problems. The most commonly used scale to clinically measure disease progression in MS is the Expanded Disability Status Scale (EDSS) introduced by Kurtzke [92]. This scale extends from 0 to 10, and is based on the functional systems (visual, brainstem, pyramidal, cerebella, sensory, bowel/bladder, and cerebral) in the lower scores, and on ambulation in the higher scores. The clinical quantification of disease severity is subjective and sometimes equivocal. The development of new treatments demands objective outcome measures for relatively short trials. Therefore, MR imaging has become one of the most important paraclinical tests for diagnosing MS and in monitoring disease progression in MS [100].

6.2 Frameworks for MR Image Analysis in MS

In any MR image analysis process, several image processing techniques are utilized—from image acquisition and tomographic reconstruction, through image enhancement, filtering, segmentation, and registration to qualitative and quantitative analysis. Since image segmentation is a very crucial step in image analysis, in this chapter, we shall focus mainly on segmentation techniques as applied to MS, the quantitative analysis approaches, and the evaluation of the methods.

In brain (and MS) MRI analysis, usually the first step is to segment the brain parenchyma (BP) and remove the extrameningeal tissues and bone. Also, in many situations, separating the gray matter (GM) and white matter (WM) is desirable (to investigate hypotheses related to these tissues), as well as separating normal and abnormal tissues such as normal appearing WM (NAWM), abnormal

normal appearing WM (ANAWM), and lesions (LS). To be able to utilize the complementary information from multiprotocol acquisitions, or for longitudinal image analysis, the images need to be in spatial correspondence. Image registration is used to achieve this correspondence. (See Chapter 5 for a task-specific comparison of image registration methods.) Intra-subject image registration is used for matching multiple acquisitions for the same study or for matching longitudinal acquisitions. Inter-subject registration is used when creating atlases (of "average" anatomy), or when using the atlases in individual cases (e.g., for matching the atlas to the study in question and classifying the tissues based on the pre-determined classification of the atlas).

Image segmentation is the process of identifying and delineating objects in images. Literature on general image segmentation spans over four decades and on tomographic image segmentation nearly 25 years. Although the MR image segmentation activity started as soon as MR images started becoming available, most of the developments are relatively recent [13]. Image segmentation approaches may be broadly classified into two groups—region-based and boundary-based. In the former, the result is a description of the region occupied by the object in the image. In the latter, the outcome is a description of the boundary of the object in the image. In both groups, the output description may be hard (crisp) or fuzzy. In the hard case, the output is a (hard) set of image elements for region-based approaches and of boundary elements for boundary-based approaches. In the fuzzy case, the respective sets are fuzzy—that is, each element in the set has a number in the range [0, 1] associated with it that represents the degree of membership of the element in the set. The degree of human assistance required in these four groups of methods ranges from 0% to 100%. In practice, the methods utilized in clinical trials invariably require some human assistance for each study for a human reader to check the results to insure that they are correct. Because the component tissues of the brain—BP, GM, WM, NAWM, ANAWM, LS—determined from MR images are the objects of our study, description will be restricted to MR brain image segmentation of these objects. A number of papers have been published on the technical aspects of brain image segmentation alone. Computerized methods of MR imaging analysis that have been developed for commonly used imaging protocols will be reviewed, for which representative references are provided.

6.3 T2-weighted Imaging

The sensitivity of T2-weighting in the detection of MS has long been recognized [100, 165]. MS lesions appear hyperintense in T2-weighted images. The clinical course of MS is mainly defined by the "baseline" disability of the patient, also called the disease "burden". The extent of lesions on T2-weighting, expressed usually by the total volume of lesions (T2LV), is currently regarded as the MRI measure of this disease burden. Usually dual echo T2 and proton density (PD) images are acquired and the lesions are detected using the combined information presented in the two images. The positive effects of three currently available US Food and Drug Administration-approved injectable medications (Betaseron, Avonex, Copaxone) on T2 lesion load (T2LV) have been demonstrated [77, 79, 80, 157].

An important limitation of T2WI is its low pathological specificity [115]. All histopathological features of MS, such as inflammation, edema, demyelination, axonal loss, gliosis and remyelination are seen as hyperintense lesions on T2WI while not all of these processes are clinically relevant. Reports from several research groups have shown only weak correlations between clinical measures of disability and the traditional T2 lesion load [46, 112, 142, 178]. There are also contradictory results among these reports—for example, that T2LV increases consistently and significantly in RR MS [87, 112] and that it does not [44, 142]. The reasons for such disagreements include imaging and processing methodological differences, subjectivity of EDSS and the non-accountability of the microscopic disease. In short, there is increasing evidence that T2LV is not a robust reliable measure of disease/disability in spite of its use in clinical trials. Nevertheless, since T2WI is the most widely used protocol, and T2LV is the most generally used quantitative measure from MRI, the techniques

for quantifying T2 lesions outnumber those for other protocols. Some of the methods are dual-echo proton-density T2-weighted image (PD/T2)-specific, the most commonly used protocol, whereas other acquisition techniques are used (perhaps with slight modifications) for lesion quantification.

6.3.1 Thresholding Techniques

One of the most popular approaches is thresholding. The methods differ in the way the images are pre-processed, the thresholds are selected, and in the post-processing operations (e.g., removing false detections).

The approach of [84] uses a step-by-step procedure, where the most obvious features in the images are identified first, and used to calibrate the application of the next step, until the desired features are identified. The brain is identified on the PD images as a "large and fairly homogeneous area". This is achieved by thresholding the subimages separately. The suspected lesions are identified in T2 images in a similar way. Some false positives are rejected by "relative thresholding" (using both T2 and PD images), then the images are deformed to a geometric model (atlas) and the remaining false positives are eliminated using the atlas (anatomical knowledge).

Wicks et al. [209] select two thresholds manually. The first is at a level which best separates brain from surrounding tissues. This often fails to remove the eyes, nose, and facial tissues (mainly in the bottom slices). With slice-by-slice segmentation from top to bottom using 3D connectivity of brain regions, non brain tissues are successfully rejected. Occasionally the eyes must be identified separately. The lesions are thresholded within the brain region, and lesions of size less than 4 pixels are excluded. Outlines are manually corrected. In practice, the lesion thresholds are set low enough to capture all lesions with their complete dim edges, so the final manual editing becomes essentially eliminating false positives.

In [200], new and enlarging lesions are identified in follow-up studies by visual inspection. Hyperintense lesions on T2-weighted images are identified and each lesion is individually delineated using a supervised method. The operator specifies a seed point in the center of the lesion and the threshold, used as a stopping criterion for region-growing, is interactively adjusted by the user. If this technique fails, manual tracing is used.

A simple thresholding method called contouring is described in [148]. First, the operator selects a point near the edge of the lesion on a slice display. Then the algorithm searches for a point with local maximum gradient in the vicinity of the selected point. This new point will be considered a starting point on the contour. Finally, the isointensity contour is followed, automatically delineating the lesion. This technique is widely used in short- and long-term studies, as well as in multicenter studies by several groups. Usually two or three observers identify and mark the lesions on hard copies, working together and making decisions by consensus. The time for lesion segmentation by the contouring method is similar to that of manual outlining. However, the semiautomated method has higher reproducibility [10, 43, 67].

A segmentation method, originally developed for reconstructing resistive head models for electroencephalographic modeling purposes, has been applied to volumetric analysis of brain tissues such as WM lesions [72] and CSF (for assessing brain atrophy) in different patient groups. The method includes image enhancement (3×3 averaging filtering), amplitude thresholding, region growing, manual editing, and decision trees. For lesion segmentation, three different threshold levels are applied to capture lesions of low, medium and high intensities.

In [52] the total visible lesion volume (TLV) is computed as a measure of macroscopic changes, using the method of [209]. A small diamond shaped ROI (61 pixels, 57 mm^2) is placed in the middle of "pure" WM region in the frontal lobe in a central slice. The average T2 value as well as several texture descriptors are computed within the ROI. Texture analysis is done using the gray level cooccurance matrix method. The authors tried to correlate contrast, correlation, homogeneity, and entropy, as well as the T2 value and TLV with EDSS. Their study suggests that texture analysis is not

able to detect any subtle structural changes in NAWM, possibly due to the limited image resolution, the small size of the ROI, and the non standard image intensity.

6.3.2 Parametric/Non-parametric Classification, Clustering

In [25], each tissue is assumed to have a bivariate normal probability distribution, that is estimated by sampling the dual-echo images at a number of locations, interactively specified by the operator. A probability distribution is calculated for each tissue, using the sample intensity pairs, and a feature map is created by assigning the label of the most probable tissue for each voxel. The MR data are then segmented and displayed using the feature map. At this stage, the user may specify additional samples for misclassified regions, if necessary, and the feature map is recalculated accordingly. The last step involves a connectivity algorithm and surface generation. In [78] a modified version of this clustering technique is used. The classifier based on parametric probability distributions is replaced by a nonparametric nearest neighbor classifier.

The same approach is followed by [90], however, they first apply an anisotropic diffusion filter to reduce the noise in the images. The operator provides about 20 sample points per tissue class for training, and using the intensity correspondence of these points, a classifier with a nonparametric statistical algorithm (k-nearest neighbor, kNN) is determined. GM, WM, and CSF voxels are combined to form the intracranial cavity (ICC). The volume is eroded to break connections that might exist between the ICC and extraneous structures (e.g., the optic nerve). Supervised connectivity is then used to remove the extraneous structures, and dilation is subsequently performed to reverse the erosion of the edited ICC. To remove the holes left by vessels, which are generally classified as background (BG), a modified connectivity algorithm is used. The mask is then applied to the segmented image and all labels outside the mask are reset to the background value.

Manual outlining and the bi-feature space method [90] are used and compared in [168]. Anisotropic diffusion filter is used for noise reduction. After manual seeding (minimum 40, usually 60–100 points) for the tissue classes (GM, WM, CSF, BG, LS) on 1–3 adjacent slices, nonparametric Parzen windowing is used to create the feature map. Connectivity (with film review for seeds) is applied and manual corrections are performed.

Clarke et al. [24] compare a parametric (maximum likelihood, ML), a nonparametric statistical (kNN) classifier, and an artificial neural network (ANN) approach. The ML classifier is based on Bayes decision rule which maximizes the *a posteriori* probabilities. The tissue distributions are assumed to be multivariate normal. The algorithm uses the training samples to calculate the mean vector and covariance matrix. Then for each voxel the probabilities for membership of each class are calculated. Finally a decision is made for the classification of each voxel choosing the class with the highest probability. This is essentially the classifier used in [25]. The kNN method does not use any particular distribution, but estimates the a posteriori probabilities from the distribution of neighbors in the feature space. kNN calculates the distance from the pixel feature vector to all the training pixel feature vectors and estimates the a posteriori probabilities from the frequency of the labels (classes) of the k nearest neighbors, where k is an odd integer (usually $k = 7$ is used).

In [116], first, WM and CSF samples are taken from "pure" tissues, and kNN distances to the WM and CSF clusters are computed and stored. Each lesion is manually identified by placing one or more seed points inside the lesion. A 3×3 region around the seed point(s) is used to get estimates for the PD-T2 distribution for that particular lesion only. A kNN feature map is created then that includes WM, CSF, and each individual lesion as a separate tissue class. After classification, the lesions are delineated by connected component labeling starting from the seed points specified in the detection step.

PD-, T2-weighted, and T1-weighted images are utilized in [201]. First, the T1-weighted image is registered to the PD-T2 volume and correction for inhomogeneity and noise are made. A qualified observer identifies samples representing the tissues of interest. Finally, the kNN algorithm is utilized to create a stack of color-coded segmented images. The authors found that including the T1 images

in the classifier significantly improved the delineation of tissues compared with any combination of 2D feature maps. Different classes of lesions representing different stages of the disease were identified with the 3D classifier. In [202], the authors further extended the method by incorporating a fourth feature image (either T1-weighted Gadolinium-enhanced, or magnetization transfer) into the classifier.

A "quantitative magnetic colour imaging (QMCI)" display technique was introduced in [4]. Three feature parameters are calculated from coregistered conventional spin-echo (CSE) T1-weighted and dual-echo fast spin-echo (FSE) images. These parameters ($\mathrm{R1} = 1/\mathrm{T1}$, $\mathrm{R2} = 1/\mathrm{T2}$, and $\mathrm{N(H)}$ (proton density)) are scaled and color-coded in the RGB channels of a single fused color image. Since human perception is more sensitive to color than to greyscale differences, the hypothesis is that these color-coded images help the observers to better distinguish different tissues and abnormalities. The technique proved useful, and improved lesion detection is reported in [14].

Images acquired by a specialized MR pulse sequence called AFFIRMATIVE (Attenuation of Fluid by Fast Inversion Recovery with Magnetization Transfer Imaging with Variable Echoes) (FLAIR/MTC) and FSE are used for segmenting normal and abnormal brain tissues in [11]. After image enhancement by anisotropic diffusion filtering and RF inhomogeneity correction, 30–40 sample points are specified by the user from 3 tissue classes: brain (including WM, GM, CSF), low intensity extrameningeal tissue (EMT) (bony structures), and high intensity EMT (scalp, bone marrow). Three 2-dimensional feature maps (late- vs. early-echo FSE, late-echo FSE vs. late-echo FLAIR/MTC, late- vs. early-echo FLAIR/MTC) are created using nonparametric Parzen windowing and a modified version of the technique of [25] is used. The three maps are simultaneously used and a voxel is classified into a given tissue region if it is classified as the same tissue by at least two of the three maps (otherwise classified as low intensity EMT). After the classification, EMT voxels are removed using 2D connectivity and mask propagation across neighboring slices [11]. One further step was included in [12] by adding post-contrast images to the set. By comparing the co-registered pre- and post-contrast T1-weighted images voxel-by-voxel, potential enhancing lesions are identified.

Intensity inhomogeneities are a common source of difficulty when using intensity-based classification. In [205], a method called adaptive segmentation is described that estimates and corrects intrascan inhomogeneities and simultaneously segments the tissues in an iterative process. In the M-step of the expectation maximization (EM) algorithm, tissue intensity properties are used to estimate the bias field, while in the E-step of the algorithm, the images are corrected for the known bias fields, and the tissues are segmented. These two steps are alternated in the iteration and convergence is achieved in 5–10 iterations. For the tissue model, either parametric (multivariate Gaussian mixture model), or nonparametric probability model estimates (such as kNN or Parzen window representation) may be used in the process. The accuracy of adaptive segmentation was found to be comparable with manual segmentation, and closer to manual segmentation than supervised multivariate classification while segmenting gray and white matter.

A new automated procedure for long-term quantitative follow-up studies is reported in [89]. The intracranial cavity (ICC) is identified on dual-echo CSE brain MR images using a supervised automated procedure [90]. The follow-up scans are aligned by rigid-body transformation using points of the 3D surface of the ICC masks. The contents of the ICC in each MR scan are segmented into four tissues, using a self-adaptive statistical algorithm [205] described above. Misclassifications due to partial volume effects are corrected using a combination of morphologic operators and connectivity criteria. Finally, a connectivity detection algorithm is used to separate the tissue classified as lesions into individual entities. (4D connectivity is used to follow lesions in time). Registration, classification of the contents of the ICC, and identification of individual lesions are fully automatic. Only identification of the ICC requires operator interaction. In each MR study, the program estimated volumes for the ICC, gray matter (GM), white matter (WM), white matter lesions (LS), CSF. The reproducibility of the system was superior to that of supervised segmentation, as evidenced by the coefficient of variation: CSF supervised 45.9% vs. automated 7.7%, GM 16.0% vs. 1.4%, WM 15.7% vs. 1.3%, and LS 39.5% vs. 52.0%.

6.3.3 Miscellaneous Advanced Methods

Unlike the previous "families" of methods, each of the methods outlined below is unique. They utilize artificial neural networks, Markov random fields, histogram modeling, anatomical atlas, deformation analysis, and fuzzy connectedness principles.

In [219], the ICC is detected using a radial method and thresholding [218]. Intensity correction is done by thin-plate spline fitting to the intensity values at the selected reference points [32]. Noise reduction is performed using anisotropic diffusion [63]. Classification is done using a supervised artificial neural network (ANN)-based approach [137]. The ANN has 3 input nodes (T1, PD, T2), one hidden layer with 10 nodes, and one output layer with 5 output nodes (BG, WM, GM, CSF, LS), and is trained by error backpropagation. The training set was obtained by sampling: 4 pixel neighborhood at 3–5 sampling sites for each tissue class. Typically 1–3 iterations of the sampling-classification-visual inspection loop was necessary to achieve satisfactory results. T1 images are not explicitly registered to the PD/T2 volumes. However, postprocessing techniques were devised to eliminate those false positives resulting from the misregistration of the feature images. Other voxels classified as LS are also eliminated, if they are close to the edge of the brain or adjacent to sulcal CSF, or are less than 10 voxels in size.

Johnston et al. [81] start with scaling the images to 8-bits, then automatically masking out the nonbrain tissue and apply RF inhomogeneity correction (homomorphic filtering). Each volume (PD- and T2-weighted images) is segmented independently using a modified iterated conditional modes (ICM) algorithm in 3D. Here, the image is assumed to represent locally dependent Markov random fields. The iterative process finds a segmentation that has maximum probability, given the observed data, neighbor interaction model, and *a priori* knowledge about the tissues that is provided by sampling. The resulting partial volume solutions from the PD- and T2-weighted images are then combined and normalized. Morphological processing and thresholding are applied to obtain WM/LS mask. Then the same methodology is applied to the masked RF corrected images to obtain the mask for lesion only.

In [33], after the main anatomical regions are roughly outlined, the tissues within the ICC are segmented by thresholding. The threshold selection is based on mathematical modeling of the intensity distributions. Each of the three main tissues (WM, GM, CSF) is modeled by a log-normal distribution instead of normal distribution because it is assumed that in the overlapping ranges the tissue histograms have longer tails. The model parameters are estimated from the image histogram, then the iterative Marquardt-Levenberg curve fitting algorithm is applied to find the model that has minimal discrepancy from the histogram obtained from the data. The thresholds are selected to be the intersection points of the model distributions.

A model-based method, wherein all brain images are mapped to a standard Talairach space and averaged in this space to get probability models for various tissues, is reported in [83]. Lesions are detected guided by the model and using decision-tree voxel classification. Each MR image volume destined for the model was transformed into the 3D standardized brain-based coordinate system of Talairach using 3 rotation, 3 translation, and 3 scale parameters. Inhomogeneity within images was reduced by homomorphic filtering. BG was segmented using edge tracing on one mid-sagittal slice, which in turn generated two points on each transverse slice that can in turn be traced yielding a stack of transversal contours (for ICC boundary). The ventricle is segmented similarly. The remaining brain tissue classes (GM, WM, external CSF) are identified with a supervised Bayesian classifier. Segmentation was performed only from eye-level to just blow the top of the head and the results were verified and manually corrected where needed. The final model is created by averaging the transformed and segmented volumes. The authors compared four classifiers: 1) minimum distance classifier (minimum Euclidean distance in feature space from the center of a tissue cluster); 2) Bayesian classifier; 3) decision tree classifier; and 4) pruned decision tree classifier. Based on how the model was incorporated into the classification procedure, four experiments were evaluated: 1) non model-based, using intensity-based features; 2) model-based, using the model to confine the

search for MS lesions to plausible WM locations (WM probability $<$ 50% non-MS), using intensity-based features; 3) model-based, using the model to provide geometric features (probabilities) in addition to the intensity-based features (entire brain is searched for MS lesions); and 4) model-based, using the model to provide geometric features while confining the search for MS lesions to plausible WM locations. Intensity-based features were the dual-echo mean intensity and standard deviation about the mean from a window around each voxel (used 1, 3, and 5 pixel windows, 2D and 3D windows). Training set consisted of 300 randomly selected tissue samples.

Thirion and Calmon [177] report a technique to detect and quantify active lesions without the need of precise segmentation. First, the two successive images are registered by rigid registration, and then a nonrigid deformation field is computed between the two registered images using a 3D deformable grid technique similar to optical flow. Evolving lesions are detected and volume variations are quantified using vector field analysis. Regions presenting large deformations (evolving tissue) and large divergence (spherical shape) are detected in the displacement vector field. Thresholding and connected component analysis are used to finally extract the centers and the approximate radii of the active lesions. This is the automated step for specifying the ROI for further analysis. The volume variation (contraction/expansion) of the lesion in a certain ROI is computed by integrating the divergence within a series of concentric spheres (or a series of isointensity surfaces).

Fuzzy connectedness method for T2WI is described in Section 6.7.2.

6.4 T1-weighted Imaging with Contrast Enhancement

In MS, Gadolinium(-DTPA) enhancement on MR images is characterized by blood-brain barrier (BBB) abnormality and intense inflammation, which represent the acute stage of the evolution of MS [68, 88]. In T1WIE, Gadolinium enhancement appears as a homogenous strongly hyperintense lesion or as a ring-shaped hyperintensity at the edge of chronic reactivated lesions. Enhancement is more sensitive than either clinical examination or T2WI in detecting disease activity and potentially can separate clinical groups [114]. Delayed imaging (50–60 minutes after Gadolinium injection), triple doses of Gadolinium, and a single dose with MT saturation to suppress normal brain have all been shown to increase the number of detectable enhancing lesions [47]. (There are disadvantages to high-dose contrast studies including possibly more false lesions, more scan time and more expense.) These results suggest that enhancement is not an "all or none" phenomenon. Enhancement may precede the development of T2 hyperintensity and clinical symptoms, which suggest that the BBB abnormality may be the crucial event in the inflammatory cascade [9, 87]. Further, we know that there is microscopic manifestation of the disease in normal appearing white matter (NAWM) which is not apparent under visual examination of the images following enhancement.

6.4.1 Methods Used with T1WIE

Three classes of methods used with T2WI have been applied to T1WIE to detect and quantify enhancing lesions: thresholding, feature map, and fuzzy connectedness. A description of these methods as used with T2WI has been given in the previous section.

The supervised region-growing algorithm in [200] is also used (in addition to segment T2 lesions) to segment enhancing lesions on post-contrast T1-weighted images. So is the semiautomatic local thresholding technique [67] used by [53].

Edwards et al. [39] use the Cavalieri method to estimate the volume of brainstem, cerebellum and upper cervical spinal cord from three-dimensional MRIs acquired with an MP-RAGE sequence. Each image is overlaid with a regular grid of test points and the points hitting the object are counted by an operator using the stereology interface in the ANALYZE software package, to give the volume estimate. The volume measures were correlated with neurological disability scores.

Lee et al. [93] generated probability maps for T1 Gd-enhancing and T2 lesions. Image segmentation and analysis were performed using a model-based atlas and registration [27]. There was a significantly higher probability for T2 lesions to be found in the central relative to the peripheral white matter, although the relative distribution of T1 Gd-enhancing lesions was not significantly different between central and peripheral white matter regions [93].

Fuzzy connectedness method for T1WIE is described in Section 6.7.3.

6.5 Magnetization Transfer Imaging

MS is now believed to be a diffuse process that extensively involves the white matter and is not restricted to the focal regions of disease activity visible as "lesions" in conventional T2, PD and T1WIE images. This notion came about mainly from magnetization transfer (MT) image analysis [36], subsequently also supported by postmortem [6] revealing microscopic disease characterized by edema, cellular infiltration and demyelination in macroscopically normal appearing white matter. In MT imaging [210], two consecutive sets of images are required, one with off-resonance pre-saturation of the relatively immobile macromolecular protons and one without. A magnetization transfer ratio (MTR) image is then computed from the two images according to $(M_o - M_s)/M_o$, where M_s and M_o are pixel intensities with and without pre-saturation. Published studies [6, 36, 42, 54, 99, 210] suggest that the MTR values for white matter in normal subjects are highly reproducible within each institution with a variation of less than 2%. Several studies have demonstrated that the microscopic aspect of the disease can be characterized by the MTR values measured in different regions (WM, lesions, the periphery of lesions) in a much more specific fashion than by T2LV.

6.5.1 MTR Analysis Methods

Since the first report of the use of MTR via ROI analysis [36], the technique has been extensively applied by many research groups active in MS research [54, 156, 167, 181, 200]. However, the method had some important drawbacks, particularly the subjectivity in selecting the ROI and the attendant variability observed in different centers, the possibility of introducing bias by confining to a selected ROI and not capturing an entire tissue region. To address these limitations, a method of analyzing the MTR histogram of the whole brain has been introduced [194]. It consists of accurately segmenting the brain parenchyma from the MT images, computing the MTR image for the segmented region, and then computing the mode, peak height, the MTR values corresponding to the 25th, 50th and 75th percentiles of the histogram of this image. The MTR values corresponding to WM cluster around the mode and those for GM and diseased WM (including NAWM) fall near its lower shoulder. Analysis in this manner demonstrated that the histogram could be used to generate robust measures of disease extent over a short time course. Figure 6.2 illustrates this approach.

Using MTR ROI analysis, correlation with clinical disability was found in only one study [54], whereas the histogram method [194] indicated correlation between histogram parameters and disease duration as well as measures of neurological impairment [193]. Utilizing this method, other centers have shown its potential advantages in long-term follow up [156, 167]. In another study [199], several MR-based measures including T2LV, T1LV, T1LV/T2LV ratio and MTR histogram parameters for BP (obtained using the method of [194]) were compared for their ability to correlate with the clinical measures of disability. The study concluded that the MTR histogram parameters produced the best correlations. Another study [198] analyzed postmortem tissue sampling of unfixed brains of 17 MS patients and concluded that T1 hypointensity (see below) and MTR were strongly associated with axonal density, underlining their importance as in vivo markers of disease progression in MS.

Global MTR values for the group of normal subjects and for the group of MS subjects were characterized by 30 different measures involving simple statistics, histographic characteristics, MTR

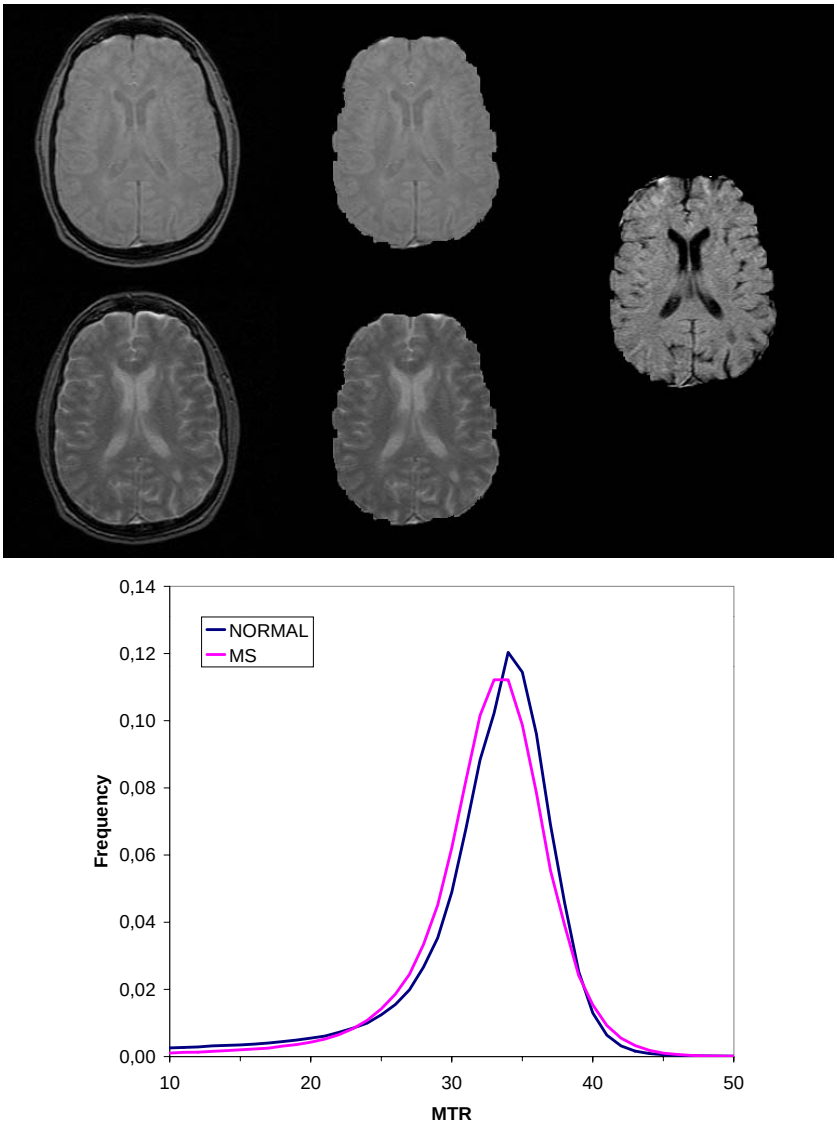


Figure 6.2: A pair of MR images with magnetization transfer (off and on, left), the corresponding pair of MT images for the brain parenchyma (middle), the MTR image for the brain parenchyma (right). The MTR histograms for a group of normal subjects and MS patients are shown on the right.

order information, and MTR range information [136].

Fuzzy connectedness method for MTI is described in Section 6.7.4.

6.6 T1-weighted Imaging

Because of the low histopathologic specificity of T2WI, while the origin of persistent deficit probably lies in the occurrence of axonal loss or extensive demyelination, more specific MR measures for these changes, such as MTR, have been studied. Another standard MRI technique that is routinely

used in MS is T1WI (SE). It often shows hypointense areas which were suggested to represent areas of axonal loss and gliosis [190]. In a recent quantitative study [200], a significant correlation was reported between increase in disability and increase in the volume of these lesions (T1LV). In another study involving a larger cohort of patients [181] including RR and SP patients, a strong correlation was found only in the SP group between increase in T1LV and increase in disability over 3 years of follow-up that was higher than the increase in T1LV. In RR patients, many T2 changes are unaccompanied by T1 changes. In contrast, in SP patients, there was a more linear relationship between accumulations of lesions in T1WI and T2WI. If hypointense lesions represent areas of severe demyelination, axonal loss or gliosis [190, 200], then the higher rate of new T2 lesions accompanied by signal loss on T1WI could reflect a deficit in available repair mechanism in SP patients. A higher "inflammatory" activity perhaps exhausts repair mechanisms leading to persistent structural loss without significant remyelination, in turn leading to persistent functional loss [150]. From this, we may conclude that the quantitative assessment of hypointense lesions on T1WI may resolve part of the clinico-radiological paradox encountered when correlating MRI lesion load with clinical disability. The observed differences between RR and SP groups probably relate to a quantitative difference in repair mechanisms.

6.6.1 Analysis Techniques for T1WI

T1 hypointense lesions (black holes) have not yet been investigated as extensively as T2 hyperintense and T1 enhancing lesions. Currently, some of the thresholding techniques used for quantifying hyperintense lesions are used with slight modifications for detecting hypointense lesions. The semi-automated contouring technique is used in [16] to obtain T1-hypointense lesion load as well as T2 lesion load. The supervised region-growing algorithm that is used for T2 and enhancing T1 lesion segmentation is also applied to black holes in T1-weighted images [2, 181, 200]. T1 hypointense lesions $> 3\,\mathrm{mm}^2$ that correspond to hyperintense regions in the T2 images are selected and quantified by a single observer using the seed-growing technique of [200]. The intraobserver lesion load variability is 5% [181]. In [2], after three observers agreed on hypointense lesion definition [181], lesion load was quantified by using the method of [209]. The segmentation technique of [72] is also used to quantify T1-weighted hypointense lesions.

6.7 Fuzzy Connectedness Methods

Objects such as WM and LS in the brain are manifest in scenes with a heterogeneity of property values (such as PD) because of the heterogeneity of material composition inherent in the object, noise, blurring, and background variations introduced by the imaging device. In spite of this graded composition of intensity values within object regions human readers perceive regions in the scene belonging to the same objects as a whole (gestalt). This property of hanging togetherness of image elements and their graded composition are both fuzzy phenomena, which should be addressed properly for effective image segmentation. While the property of graded composition has been handled in the past through fuzzy logic and/or probabilistic methods, fuzzy connectedness was the first framework that allowed capturing the idea of fuzzy hanging-togetherness via fuzzy topological principles [187]. This concept of fuzzy connected objects wherein voxels have a gradation of strength of connectedness within objects forms the basis of the methodology of [189]. An operator indicates a few points in the images by pointing to WM, GM and CSF. Each of these objects is then detected as a fuzzy connected set. The holes in the union of these objects correspond to potential lesion sites which are utilized to detect each potential lesion as a three-dimensional (3D) fuzzy connected object. These objects are presented to the operator who indicates acceptance/rejection through the click of a mouse button. The number and volume of accepted lesions is then computed and output. Based on several evaluation studies, the authors conclude that the methodology is highly reliable

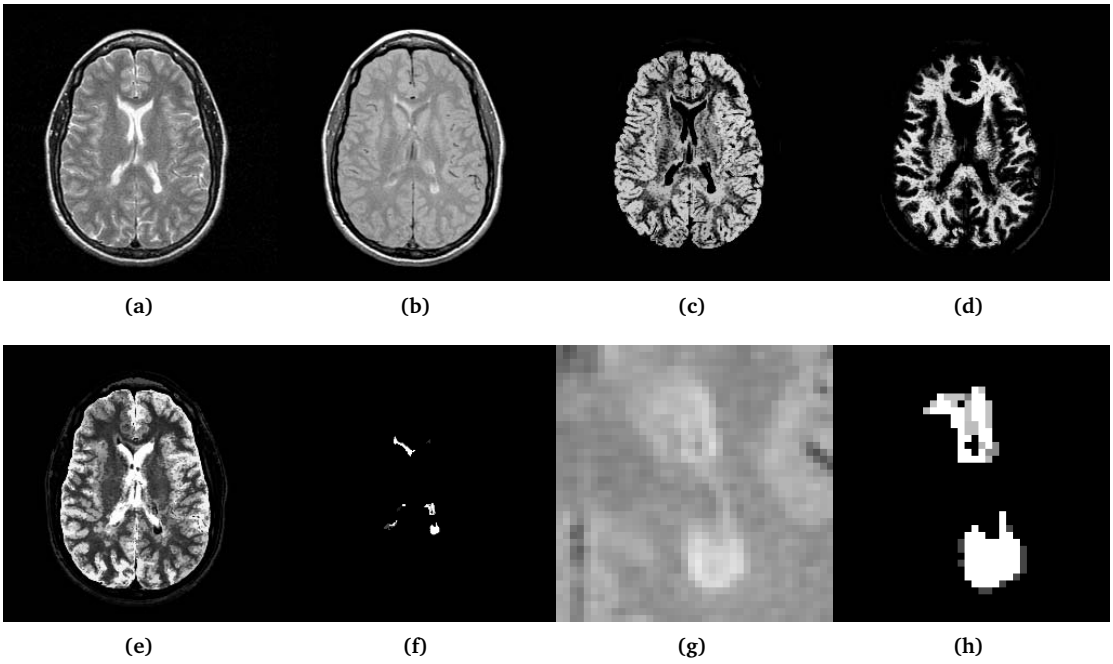


Figure 6.3: T2 (a) and PD (b) images of a MS patient. The GM (c), WM (d), CSF (e), and LS (f) objects segmented as 3D fuzzy connected objects are shown in the slices. A close up of one of the lesions (g) and its fuzzy segmentation (h) are also shown to emphasize the fuzzy nature of the lesion.

and consistent, with a coefficient of variation (due to subjective operator actions) of 0.9% (based on 20 patient studies, three operators, and two trials) for volume and a mean false-negative volume fraction of 1.3%, with a 95% confidence interval of 0%–2.8% (based on ten patient studies). The method has been used in several clinical studies subsequently [18, 20, 110, 111, 112, 146]. Figure 6.3 gives an example.

We have been developing MR image analysis methods for MS since 1993. Our approaches are guided by two premises. First, tissue regions in any anatomic object have a heterogeneous composition with or without the disease. Treating them in a binary fashion by considering a voxel to contain either 100% or 0% tissue material is unrealistic and inaccurate. The segmentation methods and the quantitative measures derived from them that we have devised take into account the heterogeneity and the fuzziness inherent in tissue regions. Second, as we have seen from the description in this chapter, different MRI protocols capture different aspects of the MS disease puzzle. We have therefore taken a multiprotocol approach wherein we utilize all major protocols currently employed in MS imaging and attempt to get a composite understanding of the most specific way of characterizing the disease through image derived parameters.

The method of fuzzy connectedness forms the essential underpinning of all segmentation algorithms utilized in our MS image analysis efforts. An outline of its principles as well as algorithms with an application-specific comparison are given in Chapter 3. In the rest of this section, we summarize our approach to analyzing the scenes for the different protocols. In the rest of this section, we describe the methods employed for extracting these parameters from the various MRI protocols, and in Section 6.8, we summarize the clinical results we have obtained to date.

Table 6.1: The MRI protocols currently used in our center and their parameters.

Pulse Sequence	Plane	TR (msec)	TE (msec)	Section Thickness (mm)	Matrix	NEX	FOV (cm)
Fast SE	Axial	2,500	18, 90	3	192	1	22
3D MT Vas TOF	Axial	106	5	5	128	1	22
SE	Axial	600	27	3	192	1	22
SE (Gd enhanced)	Axial	600	27	3	192	1	22

FOV = field of view, NEX = number of excitations, SE = spin echo, TE = echo time, 3D = three-dimensional, TOF = time of flight, TR = time to repetition.

6.7.1 Data Acquisition

The MRI protocols we were using in our patient studies are listed in Table 6.1. Every patient recruited in our trial undergoes all acquisitions listed in the table. Our image database currently consists of the following number of patient studies and 3D volume images that were processed by the methods described in this section: 690 T2WI studies, 660 T1 studies each with and without contrast, 670 MT studies, altogether 100 patients and over 4000 3D scenes. The T2 and PD images in these acquisitions are in registration since they are acquired simultaneously. Similarly the two MT images are also in registration. However, between these two sets and among other image data, registration cannot be guaranteed. Registration, when necessary, may be achived, for example, by using one of the methods described in Chapter 5.

6.7.2 Analysis of FSE T2, PD Images

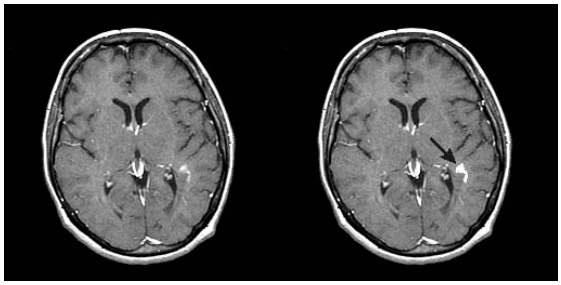
The approach here consists of the following four steps [189]. We consider here the scene to be vector-valued with two (T2 and PD) values per voxel.

1. On one slice, roughly centrally situated in the brain, an operator specifies a few voxels (seeds). This is a recognition step that uses the superior human knowledge. Points (voxels) are specified for the CSF, GM, and WM regions (and not for lesions).

2. This is a delineation step. The seed voxels are utilized to determine the fuzzy connected objects that contain them. This results first in a segmentation of WM, GM, and CSF. This knowledge is subsequently used to determine automatically a set of points (determined as holes in the union of GM and WM fuzzy objects) in each 3D lesion object that are used subsequently to delineate the lesions, each as a 3D fuzzy connected object. See Fig. 6.3.

3. This is again a recognition step taking help from a human operator. In this step, the operator accepts true lesions and rejects false lesions with a mouse button click. Each 3D lesion is displayed on one slice image passing close to the centroid of the lesion. The operator may override this mode of display and examine the lesions on all or any selected slices. All false positives are eliminated in this fashion. These are usually artifacts and choroid plexus. We found that the number of false negatives is very low in our system (see next section for details on validation). Nevertheless, the system allows in this step selecting new seed points for missed lesions. These are subsequently utilized in delineating these lesions by repeating part of Step 2 above.

4. The final step is the computation of quantitative measures from the segmented objects. These include the number of 3D lesions and their total volume (T2LV), the volume of CSF (CSFV), the volume of the brain parenchyma (BPV) which is the volume of the union of GM and WM, and normalized BPV, nBPV = BPV/(BPV + CSFV). The purpose of the last measure is to express the parenchymal volume independent of subject-to-subject variations in brain size.



(a) (b)

Figure 6.4: A T1WIE image (a) and the enhancing lesions detected via 3D fuzzy connectedness (b).

6.7.3 Analysis of T1WIE Images

Another application of the fuzzy connectedness approach for T1WIE images is described in [164]. Four neuroradiologists selected "true" lesions from the computer-detected potential lesions with a yes/no response to the program query on two different days. The enhanced lesion volume and number of enhancing lesions for each image and each observer were subsequently computed. Additional studies involving 720 slices were conducted to determine lesions that were missed by the system. The intra- and interobserver variability in the system was 0%. It took approximately 1 min of operator time per 3D study. The system output has no false positives and a mean false-negative volume of 1.3%. This method has been further evaluated in [110]. Figure 6.4 shows one example of this method.

The approach here consists of the following steps [164].

1. A conservative threshold is determined automatically from the histogram of the T1WIE 3D scene (see [164] for details). The purpose of this threshold is to select seed points automatically within the enhancing lesions. This is a recognition step.

2. The seed points are utilized to determine the fuzzy connected objects that contain them. This delineation step often results in the delineation of vessels. Taking the volume enclosed by the object as a criterion, large objects (large vessels) are automatically discarded.

3. This step is identical to Step 3 of the previous section.

4. For the accepted lesions, their total volume (T1EV) and number (nT1EV) are computed.

6.7.4 Analysis of MT Images

MTR histograms for NAWM were analyzed in [20]. The NAWM masks were obtained by fuzzy connectedness brain segmentation from FSE dual-echo images and GM/WM separation and image registration. Patel et al. [141] use fuzzy connectedness segmentation [189] to obtain the regions for MTR analysis and correlate the histogram parameters with EDSS in RR MS. Phillips et al. [146] quantitate the brain volume, CSF volume and T2 lesion load using fuzzy connectedness [164] and investigate the relationship between lesion burden and atrophy.

The steps involved in this process are as follows [194].

1. The brain is first segmented from the two MT scenes (corresponding to the off-resonance pulse on and off). The brain mask is then utilized to compute an MTR scene $\mathcal{C}^{\mathrm{BP}}_{\mathrm{MTR}} = (C^{\mathrm{BP}}_{\mathrm{MTR}}, f^{\mathrm{BP}}_{\mathrm{MTR}})$, which is such that $f^{\mathrm{BP}}_{\mathrm{MTR}}(c) = 0$ if c is not in the brain parenchymal region, otherwise $f^{\mathrm{BP}}_{\mathrm{MTR}}(c)$ will have a value as determined from the two MT scenes as described in the above section on MT imaging.

2. The PD scene is registered with the MT scene corresponding to the off-resonance pulse off [212]. Subsequently the PD and T2 scenes are redigitized to match with the MT scenes slice-to-

slice.

The redigitized scenes are segmented for WM and GM as described in Section 6.7.2. The MTR scenes $\mathcal{C}_{\mathrm{MTR}}^{\mathrm{WM}}$ and $\mathcal{C}_{\mathrm{MTR}}^{\mathrm{GM}}$ are computed subsequently.

3. For each of the MTR scenes $\mathcal{C}_{\mathrm{MTR}}^{\mathrm{BP}}$, $\mathcal{C}_{\mathrm{MTR}}^{\mathrm{WM}}$, and $\mathcal{C}_{\mathrm{MTR}}^{\mathrm{GM}}$, a histogram of the whole scene normalized to BPV is computed. From each of these histograms, the following parameters are computed: peak height normalized by BPV (nPH_x); the mode M_x; the 25th and 50th percentiles ($P25_x$, $P50_x$) and the mean m_x, where x stands for BP, WM, or GM. See Fig. 6.2.

6.8 Evaluation of the Fuzzy Connectedness Methods

In this section, we validate our fuzzy connectedness methods for MS analysis according to several criteria. We also correlate the results of these methods with clinical measures.

6.8.1 Validation

Any segmentation effort consists of (i) a theoretical or algorithmic framework, (ii) considerable engineering effort to make the framework work in the application at hand, and (iii) evaluation to establish the precision, accuracy, and efficiency of the method for the particular application.

Precision here refers to the reproducibility of the segmentation results considering all subjective actions including how the patient is positioned in the imager and any operator input required by the algorithms. *Accuracy* denotes the degree of agreement of the result with truth. *Efficiency* indicates the degree of operator help required (the degree of automation), which may be expressed as the operator time required for each study. A compromise in one factor is usually required to achieve an improvement in another. Below we summarize the precision, accuracy, and efficiency of the various analysis processes described in the previous section [56, 59, 164, 189, 194]. We evaluated our system extensively, taking into account the subjectivity that appears in steps that require operator input and in the placement of the patient in the imager.

For T2LV, the inter- and intra-operator coefficient of variation was less than 0.9%, the variations among repeated images was approximately 1.5%, the 95% confidence interval for false negative T2LV was 0%–2.8%, and the mean operator time was 10 minutes per study (on a Sun Sparc 20 workstation, reduced to about 7 minutes with a Sun Ultra Sparc workstation). For total volume of lesions on enhanecd T1-weighted images (T1ELV), there were no inter- and intra-operator variations, the false-negative rate was 1.3%, and the mean operator time was 2 minutes. For BPV, the inter- and intra-operator coefficient of variation was less than 0.5%, including repeated images, the false-negative rate was less than 1%, and the mean operator time was 5 minutes. For MTR, there was more than 99% correlation for MTR histogram parameters in repeated experiments, and the mean operator time was 10 minutes.

Any segmentation method can and will go wrong in practice. In completely automatic methods, there is a question of how to recognize and delineate false-positive and false-negative findings. By design, our system is not automatic, to facilitate taking recognition help from the user to minimize false-positive and false-negative findings. This help is obtained as efficaciously as possible, to make the system efficient and practical. We continue to improve the efficiency of our system without sacrificing precision and accuracy.

6.8.2 Clinical Correlation

In this section we summarize the results we obtained to date in analyzing the images in our database and in conducting clinical correlation analysis with the EDSS scores. For T2WI [18, 51, 56, 59, 111, 112, 113, 146], the strongest correlations were seen in our experiments for T2LV with nBPV (-0.73, $p < 0.001$) in the RR group and with $\mathrm{nPH}_{\mathrm{BP}}$ (-0.76, $p < 0.001$) and T1ELV (0.66, $p < 0.001$) both

in the RR group. T2LV also showed a good correlation with nBPV (-0.66, $p < 0.02$) in patients with SP disease. In both groups, the correlation of T2LV with EDSS scores was not significant. Generally, we found CSFV to be significantly larger in MS patients and nBPV to be smaller than in age-matched normal control subjects ($p < 0.005$). A good correlation was seen between T2LV and CSFV (0.73, $p < 0.001$).

In comparing patients with and without U-fiber lesions, we found a significant ($p < 0.05$) difference in three of the 11 neuropsychological test scores, which suggests that these lesions may lead to neuropsychological impairment. By analyzing the lesions in GM, we found that cortical GM lesions comprised about 5.7% of the total T2LV, while deep GM lesions comprised about 4.6%. No significant correlation was found between these volumes and clinical disability. The parameter computed from T2- and PD-weighted scenes that had by far the strongest correlation with disability was nBPV (-0.69, $p < 0.004$) for the patients with SP disease, which suggests that brain atrophy may be important for characterizing disease via T2-weighted imaging.

We also analyzed clinical correlation with Gadolinium-enhanced T1-weighted imaging [58, 59, 110, 112, 113]. We generally found that T1ELV was greater in RR than in SP disease ($p < 0.05$). In the RR group only, T1ELV showed good correlation with nBPV (-0.51, $p < 0.009$) and with T2LV. Compared with RR patients receiving placebo, a group of RR patients receiving Copaxone had a significant decrease in the number ($p = 0.05$) and volume ($p = 0.01$) of enhancing lesions at the end of 24 months [58]. There were no differences in T2LV between the two groups. The loss of brain tissue was significantly less ($p = 0.02$) in the group receiving Copaxone than in the placebo group.

With MT imaging [112, 113, 141, 146, 192], the $\mathrm{nPH_{BP}}$ for the MS study group (both RR and SP disease) was significantly lower than in the control group ($p < 0.0005$), as were M_{BP} ($p < 0.05$), $P25_{\mathrm{BP}}$ ($p < 0.05$), and $P50_{\mathrm{BP}}$ ($p < 0.05$). In particular, $\mathrm{nPH_{BP}}$, showed the best correlation with EDSS score for the RR group (-0.44, $p < 0.02$). This parameter correlated strongly with nBPV in both the RR (0.69, $p < 0.001$) and SP (0.87, $p < 0.001$) groups and with CSFV (-0.828, $p < 0.0001$) when the two disease groups were combined. MTR parameters also correlated with disease duration ($p < 0.01$).

Individual neuropsychological tests correlated with MTR measures ($p < 0.001$). The unnormalized histogram peak height differed ($p < 0.05$) among severely impaired, moderately impaired, and normal patients. In a serial MRI study of the RR group, $\mathrm{nPH_{BP}}$ demonstrated a subtle but significant ($p < 0.05$) decline over time, but no significant changes in EDSS were noted. MWM and mWM were significantly lower ($p < 0.05$) in RR MS patients than in normal controls [20]. In longitudinal studies, both MWM and mWM shifted in the direction of normal with increasing disease duration. Up to 44% of new lesions identified on later studies were demonstrated to have originated in the WM region, which had been identified as abnormal on the basis of MTR criteria. As to the GM, M_{GM} and m_{GM} were significantly lower ($p < 0.01$) in patients with RR MS than in normal controls [55], and $\mathrm{nPH_{GM}}$ inversely correlated with EDSS (0.65, $p < 0.01$).

By far the strongest correlations with the clinical measures were demonstrated by the MTR histogram parameters obtained for the various segmented tissue regions emphasizing the importance of considering the microscopic/diffuse nature of the disease in the individual tissue regions. (Normalized) brain parenchymal volume also demonstrated a strong correlation with the clinical measures indicating that brain atrophy is an important indicator of the disease.

6.9 Discussion and Concluding Remarks

The current difficulties in lesion segmentation and MS image analysis are well described in the review article [45] which are valid even today. "The clarity with which multiple sclerosis lesions are seen on typical MRIs belies the difficulties faced in fully automating lesion detection and delineation. The signal intensity is on an arbitrary scale: the intensity of lesions in one MRI may be very

different to that seen on another and the intensity of individual lesions is highly variable. These variations may be due to pathological heterogeneity of lesions or to partial volume effects." The graded composition of lesions resulting from pathological heterogeneity and partial volume effects are not accurately accounted for by binary delineation methods. This insistence on producing binary results removes the flexibility of handling the whole spectrum of lesions (large, small, strong, sharp, very blurred, weak, subtle, diffuse, microscopic) and consequently methodological compromises are made, such as limiting the analysis to a few slices rather than extending to the entire brain volume. Fuzzy image processing frameworks become essential to handle material heterogeneity, noise, background intensity variation, fuzziness inherent in images, and how image elements hang together in image space and feature space in the presence of this fuzziness.

6.9.1 Comparative Assessment

From the published literature, it is difficult to comparatively assess segmentation algorithms for the following reasons.

(1) Since the engineering effort required in making a methodology work in a given application is very application-specific, there is no guarantee that a methodology with an established (acceptable) precision, accuracy and efficiency will yield the same precision, accuracy, and efficiency, or even that it will work at all, in a different application situation.

(2) Most frameworks and/or their implementations are quite sensitive to small changes in the MRI protocol. This may also cause problems for data coming from different MR scanners even when identical or similar protocols are used since considerable variations in image properties often exist for such data. Often, only "good behaving" part of the brain is considered (from the eye level to just below the top of the head) because the lower brain structures (such as brain stem and cerebellum) are more affected by partial volume averaging and field inhomogeneity artifacts than the "good" parts. Most methods have not been tested on 100's of studies and we suspect that they would require per-study adjustment/training of segmentation parameters to account for scanner-dependent intensity variations.

(3) Validation is not always provided, or done only for a few slices or for selected datasets. Measures used for precision and accuracy vary considerably among published papers. In a few cases (especially for segmentation of normal tissues) physical phantoms are used as the basis of truth. For real data, the absolute truth is not known, therefore, some surrogate is used for testing accuracy of a particular method. This surrogate is almost always manually outlined segmentation.

Precision of a method is usually measured by intra- and inter-scan and -rater coefficients of variation or by different similarity measures (such as overlap or kappa statistics). Many papers report on inter- and intra-operator variabilities in volume caused by subjective human input required in segmentation, and variations due to patient positioning in the scanner are not commonly assessed. Using volume as a variate is often misleading because two sets of voxels S1 and S2 obtained as segmentations of the same object in two repeat trials may yield the same volume, yet the difference sets S1–S2 and S2–S1 may have non negligible volumes.

Efficiency is rarely measured or reported and it is difficult to assess the practical viability of the methods in processing 100's of data sets from most published segmentation papers. Efficiency evaluation often (if given at all) includes "running time" for the programs or some overall processing times based on measurements for a few datasets. Efficiency as human interaction necessary is assessed only in a few studies.

6.9.2 Automation, Failure, User Assistance

Any segmentation method can go wrong, and therefore will, if enough studies are analyzed in a routine clinical trial setting. Although more sophisticated methods usually require less human interaction than others, with no exception, they all need some operator assistance. Even if they

do not require "training" samples, or interactive guidance, some sort of verification of the output becomes necessary. Therefore, for quality assurance, it is important to have a knowledgeable human operator within the processing loop. Complete automation may therefore be an elusive goal, perhaps reachable after only a great deal of experience within the same imaging modality, protocol, and application setting. Human interaction should be considered within the design framework itself of the segmentation algorithms with a research goal of minimizing this interaction as much as possible.

6.9.3 Standardized MR Intensity-based Analysis

A major difficulty with the MRI techniques for most protocols has been that image intensities do not have a fixed meaning, even within the same protocol, for the same body region, for images obtained on the same scanner, for the same patient. See Figs. 2.1 and 2.7. This poses problems in finding the proper window setting for image display, and more importantly, in image segmentation and analysis. Any segmentation method has parameters. Without the same protocol-specific intensity meaning, setting values for the parameters in a patient-study-independent fashion becomes very difficult. This results in compromising precision, accuracy, and efficiency of segmentation. For example, per study adjustment of parameters may be necessary to handle these scanner-dependent variations in intensity, which will affect efficiency and may also lower precision. Our recently developed image processing method [128, 135] for MRI intensity standardization is described in Chapter 2. It may enable us to overcome the scanner-dependent intensity variations that pose such severe challenges to image segmentation methods. We are in the process of utilizing this method in modifying our system to improve the efficiency of the various processes of quantification without sacrificing precision and accuracy.

All objects—normal and pathological—have a heterogeneous tissue composition. This combined with the blurring and noise introduced by the imaging device makes tissue regions have a heterogeneity of intensities. We believe that this heterogeneity has tissue-specific information and is useful in characterizing disease stage and severity. Such a characterization becomes impossible when there is scanner-dependent intensity variation. Standardization may permit us to treat other protocols (T2, PD, T1WIE, T1WI) also in the same tissue-specific way as MT imaging, as illustrated with MTR analysis where MTR values have tissue-specific meaning. More importantly, flat measures such as volume of tissue regions ignore heterogeneity and may lose important disease specific information. We should really consider volume distributions, that is, intensity histograms within segmented tissue regions, as demonstrated by MTR analysis, for understanding subtle disease processes [60].

6.9.4 Segmentation "Workshop"

In MS (and other neurological applications), a variety of MRI protocols are utilized as we examined in this chapter. The actual imaging parameters used in these protocols vary among institutions. In spite of the numerous brain MR image segmentation methods developed during the past 15 years, none of them is capable of handling variations within the same protocol, and much less, the variations among protocols. What we need is a segmentation "workshop" wherein a protocol-specific segmentation method can be quickly fabricated. In Chapter 4, we described one such workshop that combines the fuzzy connectedness framework with MRI intensity scale standardization. The application context in Chapter 4 is brain segmentation. For MS, we are in the process of extending this workshop to come up with an MS specific segmentation tool.

Bibliography

[1] G. B. Aboutanos, J. Nikanne, N. Watkins, and B. M. Dawant. Model creation and deformation for the automatic segmentation of the brain in MR images. *IEEE Transactions on Bio-medical Engineering*, 46(11):1346–1356, November 1999.

[2] H. P. Adams, S. Wagner, D. F. Sobel, L. S. Slivka, J. C. Sipe, J. S. Romine, E. Beutler, and J. A. Koziol. Hypointense and hyperintense lesions on magnetic resonance imaging in secondary-progressive MS patients. *European Neurology*, 42(1):52–63, July 1999.

[3] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows : Theory, Algorithms, and Applications*, chapters 4, 5, A. Prentice-Hall, Englewood Cliffs, NJ, February 1993.

[4] B. Alfano, A. Brunetti, M. Arpaia, E. M. Covelli, and A. Ciarmiello. Multiparametric display of spin echo data from MR brain studies. *Journal of Magnetic Resonance Imaging*, 5(2):217–225, March–April 1995.

[5] B. Alfano, A. Brunetti, E. M. Covelli, M. Quarantelli, M. R. Panico, A. Ciarmiello, and M. Salvatore. Unsupervised, automated segmentation of the normal brain using a multispectral relaxometric magnetic resonance approach. *Magnetic Resonance in Medicine*, 37(1):84–93, January 1997.

[6] I. V. Allen and S. R. McKeown. A histological, histochemical and biochemical study of the macroscopically normal white matter in multiple sclerosis. *Journal of the Neurological Sciences*, 41(1):81–91, March 1979.

[7] M. S. Atkins and B. T. Mackiewich. Fully automatic segmentation of the brain in MRI. *IEEE Transactions on Medical Imaging*, 17(1):98–107, February 1998.

[8] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46(1):1–21, April 1989.

[9] H. J. Barratt, D. Miller, and P. Rudge. The site of the lesion causing deafness in multiple sclerosis. *Scandinavian Auditology*, 17(2):67–71, 1988.

[10] S. Bastianello, A. Bozzao, A. Paolillo, E. Giugni, C. Gasperini, T. Koudriavtseva, E. Millefiorini, M. A. Horsfield, C. Colonnese, D. Toni, M. Fiorelli, C. Pozzilli, and L. Bozzao. Fast spin-echo and fast fluid-attenuated inversion-recovery versus conventional spin-echo sequences for MR quantification of multiple sclerosis lesion. *AJNR, American Journal of Neuroradiology*, 18(4): 699–704, April 1997.

[11] B. J. Bedell and P. A. Narayana. Automatic removal of extrameningeal tissues from MR images of human brain. *Journal of Magnetic Resonance Imaging*, 6(6):939–943, November–December 1996.

[12] B. J. Bedell and P. A. Narayana. Automatic segmentation of gadolinium-enhanced multiple sclerosis lesions. *Magnetic Resonance in Medicine*, 39(6):935–940, June 1998.

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

[14] A. Brunetti, G. Tedeschi, A. Di Costanzo, E. M. Covelli, L. Aloj, S. Bonavita, A. Ciarmiello, B. Alfano, and M. Salvatore. White matter lesion detection in multiple sclerosis: Improved interobserver concordance with multispectral MRI display. *Journal of Neurology*, 244(9):586–590, September 1997.

[15] G. Bueno, O. Musse, F. Heitz, and J. P. Armspach. Three-dimensional segmentation of anatomical structures in MR images on large data bases. *Magnetic Resonance Imaging*, 19(1):73–88, January 2001.

[16] S. J. Camp, V. L. Stevenson, A. J. Thompson, D. H. Miller, C. Borras, S. Auriacombe, B. Brochet, M. Falautano, M. Filippi, L. Hérissé-Dulo, X. Montalban, E. Parrcira, C. H. Polman, J. De Sa, and D. W. Langdon. Cognitive function in primary progressive and transitional progressive multiple sclerosis: A controlled study with MRI correlates. *Brain*, 122(7):1341–1348, July 1999.

[17] B. M. Carvalho, C. J. Gau, G. T. Herman, and T. Y. Kong. Algorithms for fuzzy segmentation. *Pattern Analysis & Applications*, 2(1):73–81, 1999.

[18] I. Catalaa, J. C. Fulton, X. Zhang, J. K. Udupa, D. Kolson, M. Grossman, L. Wei, J. C. McGowan, M. Polansky, and R. I. Grossman. MR imaging quantitation of gray matter involvement in multiple sclerosis and its correlation with disability measures and neurocognitive testing. *AJNR, American Journal of Neuroradiology*, 20(9):1613–1618, October 1999.

[19] I. Catalaa, R. I. Grossman, D. L. Kolson, L. G. Nyúl, L. Wei, J. K. Udupa, M. Polansky, and J. C. McGowan. Magnetization transfer histogram analysis of segmented normal-appearing white matter in multiple sclerosis. In *ISMRM Seventh Meeting Proceedings, 22-28 May 1999, Philadelphia, PA*, page 957, May 1999.

[20] I. Catalaa, R. I. Grossman, D. L. Kolson, J. K. Udupa, L. G. Nyúl, L. Wei, X. Zhang, M. Polansky, L. J. Mannon, and J. C. McGowan. Multiple sclerosis: Magnetization transfer histogram analysis of segmented normal-appearing white matter. *Radiology*, 216(2):351–355, August 2000.

[21] J. M. Charcot. *Lectures on the Diseases of the Nervous System*. The New Sydenham Society, London, 1877.

[22] M. C. Clark, L. O. Hall, D. B. Goldgof, L. P. Clarke, R. P. Velthuizen, and M. S. Silbiger. MRI segmentation using fuzzy clustering techniques: Integrating knowledge. *IEEE Engineering in Medicine and Biology Magazine*, 13(5):730–742, November–December 1994.

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

[24] L. P. Clarke, R. P. Velthuizen, S. Phuphanich, J. D. Schellenberg, J. A. Arrington, and M. Silbiger. MRI: Stability of three supervised segmentation techniques. *Magnetic Resonance Imaging*, 11(1):95–106, 1993.

[25] H. E. Cline, W. E. Lorensen, R. Kikinis, and F. Jolesz. Three-dimensional segmentation of MR images of the head using probability and connectivity. *Journal of Computer Assisted Tomography*, 14(6):1037–1045, November–December 1990.

[26] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modal image registration based on information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging, 14th International Conference, Ile de Berder, France, June 1995*, volume 3 of *Computational Imaging and Vision*, pages 263–274, Dordrecht, June 1995. Kluwer Academic Publishers.

[27] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, 18(2):192–205, March–April 1994.

[28] D. L. Collins, T. M. Peters, and A. C. Evans. Automated 3D nonlinear image deformation procedure for determination of gross morphometric variability in human brain. In R. A. Robb, editor, *Visualization in Biomedical Computing 1994*, volume 2359 of *Proceedings of SPIE*, pages 180–190, September 1994.

[29] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models — Their training and applications. *Computer Vision and Image Understanding : CVIU*, 61(1):38–59, January 1995.

[30] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*, chapters 7, 12, 16, 20, 21, 25. The MIT Press, Cambridge, MA, June 1990.

[31] C. Davatzikos and N. Bran. Using a deformable surface model to obtain a shape representation of the cortex. *IEEE Transactions on Medical Imaging*, 15(6):785–795, December 1996.

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

[33] C. DeCarli, J. Maisog, D. G. M. Murphy, D. Teichberg, S. I. Rapoport, and B. Horowitz. Method for quantification of brain, ventricular, and subarachnoid CSF volumes from MR images. *Journal of Computer Assisted Tomography*, 16(2):274–284, March–April 1992.

[34] R. B. Dial. Algorithm 360: Shortest-path forest with topological ordering. *Communications of the ACM*, 12(11):632–633, November 1969.

[35] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[36] V. Dousset, R. I. Grossman, K. N. Ramer, M. D. Schnall, L. H. Young, F. Gonzalez-Scarano, E. Lavi, and J. A. Cohen. Experimental allergic encephalomyelitis and multiple sclerosis: Lesion characterization with magnetization transfer imaging. *Radiology*, 182(2):483–491, February 1992.

[37] N. Duta and M. Sonka. Segmentation and interpretation of MR brain images: An improved active shape model. *IEEE Transactions on Medical Imaging*, 17(6):1049–1062, December 1998.

[38] W. A. Edelstein, P. A. Bottomley, and L. M. Pfeifer. A signal-to-noise calibration procedure for NMR imaging systems. *Medical Physics*, 11(2):180–185, March–April 1984.

[39] S. G. M. Edwards, Q. Y. Gond, C. Liu, M. E. Zvartau, T. Jaspan, N. Roberts, and L. D. Blumhardt. Infratentorial atrophy on magnetic resonance imaging and disability in multiple sclerosis. *Brain*, 122(2):291–301, February 1999.

[40] B. J. Erickson and R. T. Avula. An algorithm for automatic segmentation and classification of magnetic resonance brain images. *Journal of Digital Imaging*, 11(2):74–82, May 1998.

[41] A. C. Evans, W. Dai, D. L. Collins, P. Neelin, and S. Marrett. Warping of a computerized 3-D atlas to match brain image volumes for quantitative neuroanatomical and functional analysis. In M. H. Loew, editor, *Medical Image Processing V: Image Processing*, volume 1445 of *Proceedings of SPIE*, pages 236–246, June 1991.

[42] M. Filippi, A. Campi, V. Dousset, C. Baratti, V. Matinelli, N. Canal, G. Scotti, and G. Comi. A magnetization transfer imaging study of normal-appearing white matter in multiple sclerosis. *Neurology*, 45(3):478–482, March 1995.

[43] M. Filippi, M. L. Gawne-Cain, C. Gasperini, J. H. van Waesberghe, J. Grimaud, F. Barkhof, M. P. Sormani, and D. H. Miller. Effect of training and different measurement strategies on the reproducibility of brain MRI lesion load measurements in multiple sclerosis. *Neurology*, 50(1):238–244, January 1998.

[44] M. Filippi, M. A. Horsfield, S. P. Morrissey, D. G. MacManus, P. Rudge, W. I. McDonald, and D. H. Miller. Quantitative brain MRI lesion load predicts the course of clinically isolated syndromes suggestive of multiple sclerosis. *Neurology*, 44(4):433–435, April 1994.

[45] M. Filippi, M. A. Horsfield, P. S. Tofts, F. Barkhof, A. J. Thompson, and D. H. Miller. Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. *Brain*, 118(6):1601–1612, December 1995.

[46] M. Filippi, D. W. Paty, L. Kappos, F. Barkhof, D. A. Compston, A. J. Thompson, G. J. Zhao, C. M. Wiles, W. I. McDonald, and D. H. Miller. Correlations between changes in disability and T2-weighted brain MRI activity in multiple sclerosis: A follow-up study. *Neurology*, 45(2): 255–260, February 1995.

[47] M. Filippi, T. Yousry, A. Campi, C. Kandziora, B. Colombo, R. Voltz, V. Martinelli, S. Spuler, S. Bressi, G. Scotti, and G. Comi. Comparison of triple dose versus standard dose gadolinium-DTPA for detection of MRI enhancing lesions in patients with MS. *Neurology*, 46(2):379–384, February 1996.

[48] J. A. Frank, editor. *Advances in Multiple Sclerosis*, volume 10(4) of *Neuroimaging Clinics of North America*. W. B. Saunders Company, Philadelphia, PA, November 2000.

[49] M. L. Fredman and R. E. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM*, 34(3):596–615, July 1987.

[50] P. A. Freeborough, N. C. Fox, and R. I. Kitney. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Computer Methods and Programs in Biomedicine*, 53(1): 15–25, May 1997.

[51] J. C. Fulton, R. I. Grossman, J. K. Udupa, L. J. Mannon, M. Grossman, L. Wei, M. Polansky, and D. L. Kolson. MR lesion load and cognitive function in patients with relapsing-remitting multiple sclerosis. *AJNR, American Journal of Neuroradiology*, 20(10):1951–1955, November–December 1999.

[52] C. Gasperini, M. A. Horsfield, J. W. Thorpe, D. Kidd, G. J. Barker, P. S. Tofts, D. G. MacManus, A. J. Thompson, D. H. Miller, and W. I. McDonald. Macroscopic and microscopic assessments of disease burden by MRI in multiple sclerosis: Relationship to clinical parameters. *Journal of Magnetic Resonance Imaging*, 6(4):580–584, July–August 1996.

[53] C. Gasperini, C. Pozzilli, S. Bastianello, E. Giugni, M. A. Horsfield, T. Koudriavtseva, S. Galgani, A. Paolillo, S. Haggiag, E. Millefiorini, and C. Fieschi. Interferon-β-1a in relapsing-remitting multiple sclerosis: Effect on hypointense lesion volume on T1 weighted images. *Journal of Neurology, Neurosurgery and Psychiatry*, 67(5):579–584, November 1999.

[54] A. Gass, G. J. Barker, D. Kidd, J. W. Thorpe, D. MacManus, A. Brennan, P. S. Tofts, A. J. Thompson, W. I. McDonald, and D. H. Miller. Correlation of magnetization transfer ratio with clinical disability in multiple sclerosis. *Annals of Neurology*, 36(1):62–67, July 1994.

[55] Y. Ge, R. I. Grossman, J. K. Udupa, J. S. Babb, D. L. Kolson, and J. C. McGowan. Magnetization transfer ratio histogram analysis of gray matter in relapsing-remitting multiple sclerosis. *AJNR, American Journal of Neuroradiology*, 22(3):470–475, March 2001.

[56] Y. Ge, R. I. Grossman, J. K. Udupa, J. S. Babb, L. G. Nyúl, and D. L. Kolson. Brain atrophy in relapsing-remitting multiple sclerosis: Fractional volumetric analysis of gray matter and white matter. *Radiology*, 220(3):606–610, September 2001.

[57] Y. Ge, R. I. Grossman, J. K. Udupa, J. S. Babb, L. G. Nyúl, and J. C. McGowan. Tissue characterization in relapsing-remitting and secondary-progressive MS via magnetization transfer ratio. In *ISMRM Eighth Meeting Proceedings, 1-7 April 2000, Denver, CO*, volume 2, page 1189, April 2000.

[58] Y. Ge, R. I. Grossman, J. K. Udupa, J. Fulton, C. S. Constantinescu, F. Gonzales-Scarano, J. S. Babb, L. J. Mannon, D. L. Kolson, and J. A. Cohen. Glatiramer acetate (Copaxone) treatment in relapsing-remitting MS: Quantitative MR assessment. *Neurology*, 54(4):813–817, February 22, 2000.

[59] Y. Ge, R. I. Grossman, J. K. Udupa, L. Wei, L. J. Mannon, M. Polansky, and D. L. Kolson. Brain atrophy in relapsing-remitting multiple sclerosis and secondary progressive multiple sclerosis: Longitudinal quantitative analysis. *Radiology*, 214(3):665–670, March 2000.

[60] Y. Ge, J. K. Udupa, L. G. Nyúl, L. Wei, and R. I. Grossman. Numerical tissue characterization in MS via standardization of the MR image intensity scale. *Journal of Magnetic Resonance Imaging*, 12(5):715–721, November 2000.

[61] Y. Ge, J. K. Udupa, L. G. Nyúl, L. Wei, and R. I. Grossman. Numerical tissue characterization in MS via standardization of the MR image intensity scale. In *ISMRM Eighth Meeting Proceedings, 1-7 April 2000, Denver, CO*, volume 1, page 579, April 2000.

[62] J. C. Gee, M. Reivich, and R. Bajcsy. Elastically deforming 3D atlas to match anatomical brain images. *Journal of Computer Assisted Tomography*, 17(2):225–236, March–April 1993.

[63] G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, June 1992.

[64] G. Gerig, D. Welti, C. R. Guttmann, A. C. Colchester, and G. Székely. Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data. *Medical Image Analysis*, 4(1):31–42, March 2000.

[65] A. F. Goldszal, C. Davatzikos, D. L. Pham, M. X. Yan, R. N. Bryan, and S. M. Resnick. An image-processing system for qualitative and quantitative volumetric analysis of brain images. *Journal of Computer Assisted Tomography*, 22(5):827–837, September–October 1998.

[66] L. Gong and C. A. Kulikowski. Composition of image analysis processes through object-centered hieararchical planning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):997–1009, 1995.

[67] J. Grimaud, M. Lai, P. Adeleine, L. Wang, G. J. Barker, D. L. Plummer, P. S. Tofts, W. I. McDonald, and D. H. Miller. Quantification of MRI lesion load in multiple sclerosis: A comparison of three computer-assisted techniques. *Magnetic Resonance Imaging*, 14(5):495–505, 1995.

[68] R. I. Grossman, F. Gonzalez-Scarano, S. W. Atlas, S. Galetta, and D. H. Silberberg. Multiple sclerosis: Gadolinium enhancement in MR imaging. *Radiology*, 161(3):721–725, December 1986.

[69] L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J. C. Bezdek. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *IEEE Transactions on Neural Networks*, 3(5):672–682, 1992.

[70] S. L. Hauser. Multiple sclerosis and other demyelinating diseases. In K. J. Isselbacher, E. Braunwald, J. D. Wilson, J. B. Martin, A. S. Fauci, and D. L. Kasper, editors, *Harrison's Principles of Internal Medicine*, pages 2287–2295. The McGraw-Hill Companies, New York, NY, 13th edition, December 1994.

[71] C. P. Hawkins, P. M. Munro, F. MacKenzie, J. Kesselring, P. S. Tofts, E. P. du Boulay, D. N. Landon, and W. I. McDonald. Duration and selectivity of blood-brain barrier breakdown in chronic relapsing experimental allergic encephalomyelitis studied by Gadolinium-DTPA and protein markers. *Brain*, 113(2):365–378, April 1990.

[72] T. Heinonen, P. Dastidar, H. Eskola, H. Frey, P. Ryymin, and E. Laasonen. Applicability of semi-automatic segmentation for volumetric analysis of brain lesions. *Journal of Medical Engeneering and Technololgy*, 22(4):173–178, July–August 1998.

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

[74] R. C. Herndon, J. L. Lancaster, J. N. Giedd, and P. T. Fox. Quantification of white matter and gray matter volumes from three-dimensional magnetic resonance volume studies using fuzzy classifiers. *Journal of Magnetic Resonance Imaging*, 8(5):1097–1105, September–October 1998.

[75] R. C. Herndon, J. L. Lancaster, A. W. Toga, and P. T. Fox. Quantification of white matter and gray matter volumes from T1 parametric images using fuzzy classifiers. *Journal of Magnetic Resonance Imaging*, 6(3):425–435, May–June 1996.

[76] D. L. Hill, C. Studholme, and D. J. Hawkes. Voxel similarity measures for automated image registration. In R. A. Robb, editor, *Visualization in Biomedical Computing 1994*, volume 2359 of *Proceedings of SPIE*, pages 205–216, September 1994.

[77] IFNB. Interferon beta-1b in the treatment of multiple sclerosis: Final outcome of the randomized controlled trial. The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group. *Neurology*, 45(7):1277–1285, July 1995.

[78] E. F. Jackson, P. A. Narayana, J. S. Wolinsky, and T. J. Doyle. Accuracy and reproducibility in volumetric analysis of multiple sclerosis lesions. *Journal of Computer Assisted Tomography*, 17 (2):200–205, March–April 1993.

[79] L. D. Jacobs, D. L. Cookfair, R. A. Rudick, R. M. Herndon, J. R. Richert, A. M. Salazar, J. S. Fischer, D. E. Goodkin, C. V. Granger, J. H. Simon, J. J. Alam, D. M. Bartoszak, D. N. Bourdette, J. Braiman, C. M. Brownscheidle, M. E. Coats, S. L. Cohan, D. S. Dougherty, R. P. Kinkel, M. K. Mass, F. E. Munschauer, 3rd, R. L. Priore, P. M. Pullicino, B. J. Scherokman,

B. Weinstock-Guttman, and R. H. Whitham. Intramuscular interferon β-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). *Annals of Neurology*, 39(3):285–294, March 1996.

[80] K. P. Johnson, B. R. Brooks, J. A. Cohen, C. C. Ford, J. Goldstein, R. P. Lisak, L. W. Myers, H. S. Panitch, J. W. Rose, R. B. Schiffer, T. Vollmer, L. P. Weiner, and J. S. Wolinsky. Extended use of glatiramer acetate (Copaxone) is well tolerated and maintains its clinical effect on multiple sclerosis relapse rate and degree of disability. Copolymer 1 Multiple Sclerosis Study Group. *Neurology*, 50(3):701–708, March 1998.

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

[82] P. Julin, T. Melin, C. Andersen, B. Isberg, L. Svensson, and L. O. Wahlund. Reliability of interactive three-dimensional brain volumetry using MP-RAGE magnetic resonance imaging. *Psychiatry Research: Neuroimaging*, 76(1):41–49, November 28, 1997.

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

[84] I. Kapouleas. Automatic detection of white matter lesions in magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 32(1):17–35, 1990.

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

[86] A. Kelemen, G. Székely, and G. Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *IEEE Transactions on Medical Imaging*, 18(10):828–839, October 1999.

[87] A. G. Kermode, A. J. Thompson, P. Tofts, D. G. MacManus, B. E. Kendall, D. P. Kingsley, I. F. Moseley, P. Rudge, and W. I. McDonald. Breakdown of the blood-brain barrier precedes symptoms and other MRI signs of new lesions in multiple sclerosis. Pathogenetic and clinical implications. *Brain*, 113(5):1477–1489, October 1990.

[88] A. G. Kermode, P. S. Tofts, A. J. Thompson, D. G. MacManus, P. Rudge, B. E. Kendall, D. P. Kingsley, I. F. Moseley, E. P. du Boulay, and W. I. McDonald. Heterogeneity of blood-brain barrier changes in multiple sclerosis: An MRI study with Gadolinium-DTPA enhancement. *Neurology*, 40(2):229–235, February 1990.

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

[90] R. Kikinis, M. E. Shenton, G. Gerig, J. Martin, M. Anderson, D. Metcalf, C. R. G. Guttmann, R. W. McCarley, W. Lorensen, H. Cline, and F. A. Jolesz. Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. *Journal of Magnetic Resonance Imaging*, 2 (6):619–629, November–December 1992.

[91] J. F. Kurtzke. Geographic distribution of multiple sclerosis: An update with special reference to Europe and the Mediterranean region. *Acta Neurologica Scandinavica*, 62(2):65–80, 1980.

[92] J. F. Kurtzke. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, 33(11):1444–1452, November 1983.

[93] M. A. Lee, S. Smith, J. Palace, S. Narayanan, N. Silver, L. Minicucci, M. Filippi, D. H. Miller, D. L. Arnold, and P. M. Matthews. Spatial mapping of T2 and Gadolinium-enhancing T1 lesion volumes in multiple sclerosis: Evidence for distinct mechanisms of lesion genesis? *Brain*, 122 (7):1261–1270, July 1999.

[94] T. Lei, J. K. Udupa, P. K. Saha, and D. Odhner. Artery-vein separation via MRA — An image processing approach. *IEEE Transactions on Medical Imaging*, 20(8):689–703, August 2001.

[95] L. Lemieux, G. Hagemann, K. Krakow, and F. G. Woermann. Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. *Magnetic Resonance in Medicine*, 42(1):127–135, July 1999.

[96] L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and D. G. Thomas. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Medical Physics*, 21(11):1749–1760, November 1994.

[97] K. O. Lim and A. Pfefferbaum. Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter. *Journal of Computer Assisted Tomography*, 13(4):588–593, July–August 1989.

[98] T. Lindeberg. *Scale-space Theory in Computer Vision*, volume 256 of *The Kluwer International Series in Engineering and Computer Science*. Kluwer Academic Publishers, Boston, MA, December 1993.

[99] L. A. Loevner, R. I. Grossman, J. A. Cohen, F. J. Lexa, D. Kessler, and D. L. Kolson. Microscopic disease in normal-appearing white matter on conventional MR images in patients with multiple sclerosis: Assessment with magnetization-transfer measurements. *Radiology*, 196(2): 511–515, August 1995.

[100] S. A. Lukes, L. E. Crooks, M. J. Aminoff, L. Kaufman, H. S. Panitch, C. Mills, and D. Norman. Nuclear magnetic resonance imaging in multiple sclerosis. *Annals of Neurology*, 13(6):592–601, June 1983.

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

[102] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373–386, December 1999.

[103] V. A. Magnotta, D. Heckel, N. C. Andreasen, T. Cizadlo, P. W. Corson, J. C. Ehrhardt, and W. T. Yuh. Measurement of brain structures with artificial neural networks: Two- and three-dimensional applications. *Radiology*, 211(3):781–790, June 1999.

[104] J. B. A. Maintz, P. A. van den Elsen, and M. A. Viergever. Comparison of edge-based and ridge-based registration of CT and MR brain images. *Medical Image Analysis*, 1(2):151–161, June 1996.

[105] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, March 1998.

[106] F. Masulli and A. Schenone. A fuzzy clustering based segmentation system as support to diagnosis in medical imaging. *Artificial Intelligence in Medicine*, 16(2):129–147, June 1999.

[107] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, June 1996.

[108] T. McInerney and D. Terzopoulos. Topology adaptive deformable surfaces for medical image volume segmentation. *IEEE Transactions on Medical Imaging*, 18(10):840–850, October 1999.

[109] C. R. Meyer, G. S. Leichtman, J. A. Brunberg, R. L. Wahl, and L. E. Quint. Simultaneous usage of homologous points, lines and planes for optimal, 3D, linear registration of multimodality imaging data. *IEEE Transactions on Medical Imaging*, 14(1):1–11, March 1995.

[110] Y. Miki, R. I. Grossman, J. K. Udupa, S. Samarasekera, M. A. van Buchem, B. S. Cooney, S. N. Pollack, D. L. Kolson, C. Constantinescu, M. Polansky, and L. J. Mannon. Computer-assisted quantitation of enhancing lesions in multiple sclerosis: Correlation with clinical classification. *AJNR, American Journal of Neuroradiology*, 18(4):705–710, April 1997.

[111] Y. Miki, R. I. Grossman, J. K. Udupa, L. Wei, D. L. Kolson, L. J. Mannon, and M. Grossman. Isolated U-fiber involvement in MS: Preliminary observations. *Neurology*, 50(5):1301–1306, May 1998.

[112] Y. Miki, R. I. Grossman, J. K. Udupa, L. Wei, M. Polansky, L. J. Mannon, and D. L. Kolson. Relapsing-remitting multiple sclerosis: Longitudinal analysis of MR images — Lack of correlation between changes in T2 lesion volume and clinical findings. *Radiology*, 213(2):395–399, November 1999.

[113] Y. Miki, R. I. Grossman, M. A. van Buchem, L. Wei, M. D. Phillips, U. Patel, J. C. McGowan, and D. L. Kolson. Differences between relapsing remitting and chronic progressive multiple sclerosis as determined with quantitative MR imaging. *Radiology*, 210(3):769–774, March 1999.

[114] D. H. Miller, F. Barkhof, and J. J. Nauta. Gadolinium enhancement increases the sensitivity of MRI in detecting disease activity in multiple sclerosis. *Brain*, 116(5):1077–1094, October 1993.

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

[116] J. R. Mitchell, S. J. Karlik, D. H. Lee, M. Eliasziw, G. P. Rice, and A. Fenster. The variability of manual and computer assisted quantification of multiple sclerosis lesion volumes. *Medical Physics*, 23(1):85–97, January 1996.

[117] F. B. Mohamed, S. Vinitski, S. H. Faro, C. F. Gonzalez, J. Mack, and T. Iwanaga. Optimization of tissue segmentation of brain MR images based on multispectral 3D feature maps. *Magnetic Resonance Imaging*, 17(3):304–409, April 1999.

[118] G. Moonis, J. Liu, J. K. Udupa, and D. B. Hackney. Estimation of tumor volume with fuzzy-connectedness segmentation of MR images. *AJNR, American Journal of Neuroradiology*, 23 (3):356–363, March 2002.

[119] B. Moretti, L. M. Fadili, S. Ruan, N. Bloyet, and B. Mazoyer. Phantom-based performance evaluation: Application to brain segmentation from magnetic resonance images. *Medical Image Analysis*, 4(4):303–316, December 2000.

[120] G. M. Nesbit, G. S. Forbes, B. W. Scheithauer, H. Okazaki, and M. Rodriguez. Multiple sclerosis: Histopathologic and MR and/or CT correlation in 37 cases at biopsy and three cases at autopsy. *Radiology*, 180(2):467–474, August 1991.

[121] W. L. Nowinski. Dual probabilistic classifier for three-dimensional neuroimaging from MRI data. In R. A. Robb, editor, *Visualization in Biomedical Computing 1994*, volume 2359 of *Proceedings of SPIE*, pages 373–384, September 1994.

[122] L. Nyúl and J. K. Udupa. Többdimenziós MRI képek feldolgozása. In A. Kuba, E. Máté, and K. Palágyi, editors, *Book of Abstracts, KEPAF 2002, Domaszék, Hungary, 23-25 January, 2002*, pages 96–97, January 2002.

[123] L. G. Nyúl, A. X. Falcão, and J. K. Udupa. Fuzzy-connected 3D object segmentation at interactive speeds. *Graphical Models*, 1999. submitted for publication.

[124] L. G. Nyúl, A. X. Falcão, and J. K. Udupa. Fuzzy-connected 3D image segmentation at interactive speeds. In K. M. Hanson, editor, *Medical Imaging 2000: Image Processing*, volume 3979 of *Proceedings of SPIE*, pages 212–223, June 2000.

[125] L. G. Nyúl and J. K. Udupa. On standardizing the MR image intensity scale. *Radiology*, 209 (P):581–582, November 1998.

[126] L. G. Nyúl and J. K. Udupa. An approach to standardizing the MR image intensity scale. In S. K. Mun and Y. Kim, editors, *Medical Imaging 1999: Image Display*, volume 3658 of *Proceedings of SPIE*, pages 595–603, May 1999.

[127] L. G. Nyúl and J. K. Udupa. New variants of a method of MRI scale normalization. In A. Kuba, M. Šámal, and A. Todd-Pokropek, editors, *Information Processing in Medical Imaging: 16th International Conference, IPMI'99, Proceedings*, volume 1613 of *Lecture Notes on Computer Science*, pages 490–495, Berlin, June–July 1999.

[128] L. G. Nyúl and J. K. Udupa. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine*, 42(6):1072–1081, December 1999.

[129] L. G. Nyúl and J. K. Udupa. MR image analysis in multiple sclerosis. In J. A. Frank, editor, *Advances in Multiple Sclerosis*, volume 10(4) of *Neuroimaging Clinics of North America*, pages 799–816. W. B. Saunders Company, Philadelphia, PA, November 2000.

[130] L. G. Nyúl and J. K. Udupa. Standardizing the MR image intensity scale and its applications. In T. Csendes, editor, *Book of Abstracts, Conference of PhD Students in Computer Science, July 20-23, 2000, Szeged, Hungary*, page 75, July 2000.

[131] L. G. Nyúl and J. K. Udupa. Standardizing the MR image intensity scales: Making MR intensities have tissue-specific meaning. In S. K. Mun, editor, *Medical Imaging 2000: Image Display*, volume 3976 of *Proceedings of SPIE*, pages 496–504, April 2000.

[132] L. G. Nyúl and J. K. Udupa. A protocol-independent brain MRI segmentation method. In M. Sonka and J. M. Fitzpatrick, editors, *Medical Imaging 2002: Image Processing*, volume 4684 of *Proceedings of SPIE*, pages 1588–1599, May 2002.

[133] L. G. Nyúl, J. K. Udupa, and P. K. Saha. Task-specific comparison of 3-D image registration methods. *IEEE Transactions on Medical Imaging*, 2001. submitted for publication.

[134] L. G. Nyúl, J. K. Udupa, and P. K. Saha. Task-specific comparison of 3-D image registration methods. In M. Sonka and K. M. Hanson, editors, *Medical Imaging 2001: Image Processing*, volume 4322 of *Proceedings of SPIE*, pages 1588–1598, July 2001.

[135] L. G. Nyúl, J. K. Udupa, and X. Zhang. New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–150, February 2000.

[136] J. L. Ostuni, N. D. Richert, B. K. Lewis, and J. A. Frank. Characterization of differences between multiple sclerosis and normal brain: A global magnetization transfer application. *AJNR, American Journal of Neuroradiology*, 20(3):501–507, March 1999.

[137] M. Özkan, B. M. Dawant, and R. J. Maciunas. Neural-network-based segmentation of multimodal medical images: A comparative and prospective study. *IEEE Transactions on Medical Imaging*, 12(3):534–544, September 1993.

[138] N. R. Pal, J. C. Bezdek, and E. C.-K. Tsao. Generalized clustering networks and Kohonen's self-organizing scheme. *IEEE Transactions on Neural Networks*, 4(4):549–557, July 1993.

[139] N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26 (9):1277–1294, September 1993.

[140] K. Palágyi and J. K. Udupa. Medical image registration based on fuzzy objects. In *Proceedings of the 3rd Workshop on Computational Modelling, Imaging and Visualization in Biosciences, COMBIO'96, Sopron, Hungary, 29-31 August 1996*, pages 44–48, 1996.

[141] U. J. Patel, R. I. Grossman, M. D. Phillips, J. K. Udupa, J. C. McGowan, Y. Miki, L. Wei, M. Polansky, M. A. van Buchem, and D. Kolson. Serial analysis of magnetization-transfer histograms and Expanded Disability Status Scale scores in patients with relapsing-remitting multiple sclerosis. *AJNR, American Journal of Neuroradiology*, 20(10):1946–1950, November–December 1999.

[142] D. W. Paty and D. K. Li. Interferon β-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group. *Neurology*, 43(4): 662–667, April 1993.

[143] C. A. Pelizzari, G. T. Chen, D. R. Spelbring, R. R. Weichselbaum, and C. T. Chen. Accurate three-dimensional registration of CT, PET, and/or MR images of the brain. *Journal of Computer Assisted Tomography*, 13(1):20–26, January–February 1989.

[144] G. P. Penney, J. Weese, J. A. Little, P. Demedt, D. L. G. Hill, and D. J. Hawkes. A comparison of similarity measures for use in 2-D-3-D medical image registration. *IEEE Transactions on Medical Imaging*, 17(4):586–595, August 1998.

[145] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000.

[146] M. D. Phillips, R. I. Grossman, Y. Miki, L. Wei, D. L. Kolson, M. A. van Buchem, M. Polansky, J. C. McGowan, and J. K. Udupa. Comparison of T2 lesion volume and magnetization transfer ratio histogram analysis and of atrophy and measures of lesion burden in patients with multiple sclerosis. *AJNR, American Journal of Neuroradiology*, 19(6):1055–1060, June–July 1998.

[147] S. M. Pizer, D. Eberly, D. S. Fritsch, and B. S. Morse. Zoom-invariant vision of figural shape: The mathematics of cores. *Computer Vision and Image Understanding : CVIU*, 69(1):55–71, January 1998.

[148] D. L. Plummer. Dispimage: A display and analysis tool for medical images. *Rivista di Neuroradiologia*, 5:489–495, 1992.

[149] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 2nd edition, 1992.

[150] J. W. Prineas, R. O. Barnard, T. Revesz, E. E. Kwon, L. Sharer, and E. S. Cho. Multiple sclerosis. Pathology of recurrent lesions. *Brain*, 116(3):681–693, June 1993.

[151] J. C. Rajapakse, C. DeCarli, A. McLaughlin, J. N. Giedd, A. L. Krain, S. D. Hamburger, and J. L. Rapoport. Cerebral magnetic resonance image segmentation using data fusion. *Journal of Computer Assisted Tomography*, 20(2):206–218, March–April 1996.

[152] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport. Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, 16(2):176–186, April 1997.

[153] W. E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on Medical Imaging*, 16(6):911–918, December 1997.

[154] A. L. Reiss, J. G. Hennessey, M. Rubin, L. Beach, M. T. Abrams, I. S. Warofsky, A. M. Liu, and J. M. Links. Reliability and validity of an algorithm for fuzzy tissue segmentation of MRI. *Journal of Computer Assisted Tomography*, 22(3):471–479, May–June 1998.

[155] B. L. Rice, Jr. and J. K. Udupa. Clutter-free volume rendering for magnetic resonance angiography using fuzzy connectedness. *International Journal of Imaging Systems and Technology*, 11(1):62–70, 2000.

[156] N. Richert, J. Ostuni, C. Bash, J. Duyn, H. McFarland, and J. Frank. Serial whole-brain magnetization transfer imaging in patients with relapsing-remitting multiple sclerosis at baseline and during treatment with interferon β-1b. *AJNR, American Journal of Neuroradiology*, 19 (9):1705–1713, October 1998.

[157] R. A. Rudick, D. E. Goodkin, L. D. Jacobs, D. L. Cookfair, R. M. Herndon, J. R. Richert, A. M. Salazar, J. S. Fischer, C. V. Granger, J. H. Simon, J. J. Alam, N. A. Simonian, M. K. Campion, D. M. Bartoszak, D. N. Bourdette, J. Braiman, C. M. Brownscheidle, M. E. Coats, S. L. Cohan, D. S. Dougherty, R. P. Kinkel, M. K. Mass, F. E. Munschauer, R. L. Priore, R. H. Whitham, and et al. Impact of interferon β-1a on neurologic disability in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). *Neurology*, 49(2):358–363, August 1997.

[158] V. M. Runge, A. C. Price, H. S. Kirshner, J. H. Allen, C. L. Partain, and A. E. James, Jr. Magnetic resonance imaging of multiple sclerosis: A study of pulse-technique efficacy. *AJR, American Journal of Roentgenology*, 143(5):1015–1026, November 1984.

[159] P. K. Saha and J. K. Udupa. Fuzzy connected object delineation: Axiomatic path strength definition and the case of multiple seeds. *Computer Vision and Image Understanding : CVIU*, 83(3):275–295, September 2001.

[160] P. K. Saha and J. K. Udupa. Relative fuzzy connectedness among multiple objects: Theory, algorithms, and applications in image segmentation. *Computer Vision and Image Understanding : CVIU*, 82(1):42–56, April 2001.

[161] P. K. Saha and J. K. Udupa. Scale-based diffusive image filtering preserving boundary sharpness and fine structures. *IEEE Transactions on Medical Imaging*, 20(11):1140–1155, November 2001.

[162] P. K. Saha, J. K. Udupa, E. F. Conant, D. P. Chakraborty, and D. Sullivan. Breast tissue glandularity quantification via digitized mammograms. *IEEE Transactions on Medical Imaging*, 20 (8):792–803, August 2001.

[163] P. K. Saha, J. K. Udupa, and D. Odhner. Scale-based fuzzy connected image segmentation: Theory, algorithms, and validation. *Computer Vision and Image Understanding : CVIU*, 77(2): 145–174, February 1999.

[164] S. Samarasekera, J. K. Udupa, Y. Miki, L. Wei, and R. I. Grossman. A new computer-assisted method for the quantification of enhancing lesions in multiple sclerosis. *Journal of Computer Assisted Tomography*, 21(1):145–151, January–February 1997.

[165] J. J. Sheldon, R. Siddharthan, J. Tobias, W. A. Sheremata, K. Soila, and M. Viamonte, Jr. MR imaging of multiple sclerosis: Comparison with clinical and CT examinations in 74 patients. *AJR, American Journal of Roentgenology*, 145(5):957–964, November 1985.

[166] J. Sijbers, P. Scheunders, M. Verhoye, A. van der Linden, D. van Dyck, and E. Raman. Watershed-based segmentation of 3D MR data for volume quantization. *Magnetic Resonance Imaging*, 15(6):679–688, 1997.

[167] N. C. Silver, E. Hughes, K. Birnie, P. D. Molyneux, D. G. MacManus, G. J. Barker, P. S. Tofts, and D. H. Miller. Magnetization transfer imaging provides a highly reproducible quantitative measure of brain tissue structural integrity. In *Proceedings of ISMRM 98*, page M 2162, 1998.

[168] J. H. Simon, A. Scherzinger, U. Raff, and X. Li. Computerized method of lesion volume quantitation in multiple sclerosis: Error of serial studies. *AJNR, American Journal of Neuroradiology*, 18(3):580–582, March 1997.

[169] M. E. Smith, L. A. Stone, P. S. Albert, J. A. Frank, R. Martin, M. Armstrong, H. Maloni, D. E. McFarlin, and H. F. McFarland. Clinical worsening in multiple sclerosis is associated with increased frequency and area of gadopentetate dimeglumine-enhancing magnetic resonance imaging lesions. *Annals of Neurology*, 33(5):480–489, May 1993.

[170] M. Sonka, S. K. Tadikonda, and S. M. Collins. Knowledge-based interpretation of MR brain images. *IEEE Transactions on Medical Imaging*, 15(4):443–452, August 1996.

[171] R. Stokking, K. L. Vincken, and M. A. Viergever. Automatic morphology-based brain segmentation (MBRASE) from MRI-T1 data. *NeuroImage*, 12(6):726–738, December 2000.

[172] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Multiresolution voxel similarity measures for MR-PET registration. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging, 14th International Conference, Ile de Berder, France, June 1995*, volume 3 of *Computational Imaging and Vision*, pages 287–298, Dordrecht, June 1995. Kluwer Academic Publishers.

[173] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Automated 3-D registration of MR and CT images of the head. *Medical Image Analysis*, 1(2):163–175, June 1996.

[174] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1):25–35, January 1997.

[175] J. Suckling, T. Sigmundsson, K. Greenwood, and E. T. Bullmore. A modified fuzzy clustering algorithm for operator independent brain tissue classification of dual echo MR images. *Magnetic Resonance Imaging*, 17(7):1065–1076, September 1999.

[176] R. E. Tarjan. *Data Structures and Network Algorithms*, volume 44 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, chapters 3, 7. SIAM Publications, Philadelphia, PA, 1983.

[177] J.-P. Thirion and G. Calmon. Deformation analysis to detect and quantify active lesions in 3D medical image sequences. *IEEE Transactions on Medical Imaging*, 18(5):429–441, May 1999.

[178] A. J. Thompson, A. G. Kermode, D. G. MacManus, B. E. Kendall, D. P. Kingsley, I. F. Moseley, and W. I. McDonald. Patterns of disease activity in multiple sclerosis: Clinical and magnetic resonance imaging study. *BMJ : British Medical Journal*, 300(6725):631–634, March 10, 1990.

[179] A. J. Thompson, A. G. Kermode, D. Wicks, D. G. MacManus, B. E. Kendall, D. P. Kingsley, and W. I. McDonald. Major differences in the dynamics of primary and secondary progressive multiple sclerosis. *Annals of Neurology*, 29(1):53–62, January 1991.

[180] K. D. Toennies, J. K. Udupa, G. T. Herman, I. L. Wornom, III, and S. R. Buchman. Registration of 3D objects and surfaces. *IEEE Computer Graphics and Applications*, 10(3):52–62, May–June 1990.

[181] L. Truyen, J. H. van Waesberghe, M. A. van Walderveen, B. W. van Oosten, C. H. Polman, O. R. Hommes, H. J. Ader, and F. Barkhof. Accumulation of hypointense lesions ("black holes") on T1 spin-echo MRI correlates with disease progression in multiple sclerosis. *Neurology*, 47(6): 1469–1476, December 1996.

[182] J. K. Udupa and D. C. Hemmy. Fuzzy connected object rendering. In Y. Kim and S. K. Mun, editors, *Medical Imaging 1998: Image Display*, volume 3335 of *Proceedings of SPIE*, pages 452–461, June 1998.

[183] J. K. Udupa, V. R. LaBlanc, H. Schmid, C. Imielinska, P. K. Saha, G. J. Grevera, Y. Zhuge, L. M. Currie, P. Molholt, and Y. Jin. A methodology for evaluating image segmentation algorithms. In M. Sonka and M. J. Fitzpatrick, editors, *Medical Imaging 2002: Image Processing*, volume 4684 of *Proceedings of SPIE*, pages 266–277, May 2002.

[184] J. K. Udupa, L. G. Nyúl, Y. Ge, and R. I. Grossman. Multiprotocol MR image segmentation in multiple sclerosis: Experience with over 1000 studies. In K. M. Hanson, editor, *Medical Imaging 2000: Image Processing*, volume 3979 of *Proceedings of SPIE*, pages 1017–1027, June 2000.

[185] J. K. Udupa, L. G. Nyúl, Y. Ge, and R. I. Grossman. Multiprotocol MR image segmentation in multiple sclerosis: Experience with over 1000 studies. *Academic Radiology*, 8(11):1116–1126, November 2001.

[186] J. K. Udupa, D. Odhner, S. Samarasekera, R. J. Goncalves, K. Iyer, K. Venugopal, and S. Furuie. 3DVIEWNIX: An open, transportable, multidimensional, multimodality, multiparametric imaging software system. In Y. Kim, editor, *Medical Imaging 1994: Image Capture, Formatting, and Display*, volume 2164 of *Proceedings of SPIE*, pages 58–73, May 1994.

[187] J. K. Udupa and S. Samarasekera. Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graphical Models and Image Processing*, 58 (3):246–261, May 1996.

[188] J. K. Udupa, J. Tian, D. C. Hemmy, and P. Tessier. A pentium personal computer-based craniofacial three-dimensional imaging and analysis system. *The Journal of Craniofacial Surgery*, 8(5):333–339, September 1997.

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

[190] D. Uhlenbrock and S. Sehlen. The value of T1-weighted images in the differentiation between MS, white matter lesions, and subcortical arteriosclerotic encephalopathy (SAE). *Neuradiology*, 31(3):203–212, 1989.

[191] M. Vaidyanathan, L. P. Clarke, C. Heidtman, R. P. Velthuizen, and L. O. Hall. Normal brain volume measurements using multispectral MRI segmentation. *Magnetic Resonance Imaging*, 15(1):87–97, 1997.

[192] M. A. van Buchem, R. I. Grossman, C. Armstrong, M. Polansky, Y. Miki, F. H. Heyning, M. P. Boncoeur-Martel, L. Wei, J. K. Udupa, M. Grossman, D. L. Kolson, and J. C. McGowan. Correlation of volumetric magnetization transfer imaging with clinical data in MS. *Neurology*, 50 (6):1609–1617, June 1998.

[193] M. A. van Buchem, J. C. McGowan, D. L. Kolson, M. Polansky, and R. I. Grossman. Quantitative volumetric magnetization transfer analysis in multiple sclerosis: Estimation of macroscopic and microscopic disease burden. *Magnetic Resonance in Medicine*, 36(4):632–636, October 1996.

[194] M. A. van Buchem, J. K. Udupa, J. C. McGowan, Y. Miki, F. H. Heyning, M. P. Boncoeur-Martel, D. L. Kolson, M. Polansky, and R. I. Grossman. Global volumetric estimation of disease burden in multiple sclerosis based on magnetization transfer imaging. *AJNR, American Journal of Neuroradiology*, 18(7):1287–1290, August 1997.

[195] P. A. van den Elsen, J. B. A. Maintz, E.-J. D. Pol, and M. A. Viergever. Automatic registration of CT and MR brain images using correlation of geometrical features. *IEEE Transactions on Medical Imaging*, 14(2):384–396, June 1995.

[196] P. A. van den Elsen, E.-J. D. Pol, and M. A. Viergever. Medical image matching — A review with classification. *IEEE Engineering in Medicine and Biology Magazine*, 12(1):26–39, March 1993.

[197] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, 20(8):677–688, August 2001.

[198] J. H. van Waesberghe and F. Barkhof. Magnetization transfer imaging of the spinal cord and the optic nerve in patients with multiple sclerosis. *Neurology*, 53(5 Suppl 3):S46–S48, 1999.

[199] J. H. van Waesberghe, M. A. van Buchem, M. Filippi, J. A. Castelijns, M. A. Rocca, R. van der Boom, C. H. Polman, and F. Barkhof. MR outcome parameters in multiple sclerosis: Comparison of surface-based thresholding segmentation and magnetization transfer ratio histographic analysis in relation to disability (A preliminary note). *AJNR, American Journal of Neuroradiology*, 19(10):1857–1862, November–December 1998.

[200] M. A. van Walderveen, F. Barkhof, O. R. Hommes, C. H. Polman, H. Tobi, S. T. Frequin, and J. Valk. Correlating MRI and clinical disease activity in multiple sclerosis: Relevance of hypointense lesions on short-TR/short-TE (T1-weighted) spin-echo images. *Neurology*, 45 (9):1684–1690, September 1995.

[201] S. Vinitski, C. Gonzalez, F. Mohamed, T. Iwanaga, R. L. Knobler, K. Khalili, and J. Mack. Improved intracranial lesion characterization by tissue segmentation based on a 3D feature map. *Magnetic Resonance in Medicine*, 37(3):457–469, March 1997.

[202] S. Vinitski, C. F. Gonzalez, R. Knobler, D. Andrews, T. Iwanaga, and M. Curtis. Fast tissue segmentation based on a 4D feature map in characterization of intracranial lesions. *Journal of Magnetic Resonance Imaging*, 9(6):768–776, June 1999.

[203] D. Wang, G. J. Galloway, G. I. de Zubicaray, S. E. Rose, J. B. Chalk, D. M. Doddrell, and J. Semple. A reproducible method for automated extraction of brain volumes from 3D human head MR images. *Journal of Magnetic Resonance Imaging*, 8(2):480–486, March–April 1998.

[204] S. K. Warfield, M. Kaus, F. A. Jolesz, and R. Kikinis. Adaptive, template moderated, spatially varying statistical classification. *Medical Image Analysis*, 4(1):43–55, March 2000.

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

[206] W. M. Wells, III, P. Viola, H. Atsumi, S. Nakajuma, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, March 1996.

[207] R. E. Wendt, 3rd. Automatic adjustment of contrast and brightness of magnetic resonance images. *Journal of Digital Imaging*, 7(2):95–97, May 1994.

[208] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, Jr., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawks, C. Studholme, J. B. A. Maintz, M. A. Viegever, G. Malandain, R. X. Pennec, M. E. Noz, G. Q. Maguire, Jr., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods. Comparison and evaluation of retrospective intermodality registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554–566, July–August 1997.

[209] D. A. Wicks, P. S. Tofts, D. H. Miller, G. H. du Boulay, A. Feinstein, R. P. Sacares, I. Harvey, R. Brenner, and W. I. McDonald. Volume measurement of multiple sclerosis lesions with magnetic resonance images. A preliminary study. *Neuradiology*, 34(6):475–479, 1992.

[210] S. D. Wolff and R. S. Balaban. Magnetization transfer contrast (MTC) and tissue water proton relaxation in vivo. *Magnetic Resonance in Medicine*, 10(1):135–144, April 1989.

[211] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta. Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22(1):139–152, January–February 1998.

[212] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, 17(4):536–546, July–August 1993.

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

[214] T. Yamamoto, T. Nambu, K. Miyasaka, and Y. Morita. Accurate and practical calibration of MR signal intensities by the new transmission amplitude method: Application of the numerical diagnosis to MRI. *Radiology*, 209(P):582, 1998.

[215] A. Zavaljevski, A. P. Dhawan, M. Gaskil, W. Ball, and J. D. Johnson. Multi-level adaptive segmentation of multi-parameter MR brain images. *Computerized Medical Imaging and Graphics*, 24(2):87–98, March–April 2000.

[216] Y. Zhuge, J. K. Udupa, J. Liu, P. K. Saha, and T. Iwanaga. Scale-based method for correcting background intensity variation in acquired images. In M. Sonka and M. J. Fitzpatrick, editors, *Medical Imaging 2002: Image Processing*, volume 4684 of *Proceedings of SPIE*, pages 1103–1111, May 2002.

[217] Y. Zhuge, J. K. Udupa, and P. K. Saha. Vectorial-scale-based fuzzy-connected image segmentation. In M. Sonka and M. J. Fitzpatrick, editors, *Medical Imaging 2002: Image Processing*, volume 4684 of *Proceedings of SPIE*, pages 1476–1487, May 2002.

[218] A. P. Zijdenbos, B. M. Dawant, and R. A. Margolin. Automatic detection of intracranial contours in MR images. *Computerized Medical Imaging and Graphics*, 18(1):11–23, January–February 1994.

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

Summary

This thesis deals with issues from different aspects of medical image processing—image segmentation (Chapters 3 and 4), image registration (Chapter 5), and image analysis (Chapter 6). Chapter 2 describes a preprocessing procedure that facilitates both image segmentation and analysis. Although the techniques and algorithms are described within the context of an actively pursued application, namely the analysis of brain MRI in Multiple Sclerosis, most of the methods described here are applicable to other brain diseases and other modalities as well. They even have potential in non-medical applications.

Image Processing

Image processing, whether it be applied to robotics, computer vision, or medicine, will treat imaging geometry, linear transforms, the frequency domain, digital and continuous domains, segmentation, histogram analysis, morphology, and other topics that apply to any imaging modality and any application. Image analysis, regardless of its application area, encompasses the incorporation of prior knowledge, the classification of features, the matching of models to images, the description of shape, and many of the generic problems and approaches of artificial intelligence. However, while these classic approaches to general images and to general applications are important, the special nature of medical images and medical applications requires special treatment.

Medical Image Processing

The special nature of medical images derives as much from their method of acquisition as it does from the subjects whose images are being acquired. The development of medical imaging over the past three decades has been revolutionary. With the increasing availability of relatively inexpensive computational resources, computerized techniques such as computed tomography (CT), magnetic resonance imaging (MRI), digital subtraction angiography, Doppler ultrasound imaging, and various imaging techniques based on nuclear emission (PET, SPECT, etc.) have all been valuable additions to the radiologist's arsenal of imaging tools toward ever more reliable detection and diagnosis of disease. Data sets in two, three, or more dimensions convey increasingly vast and detailed information for clinical or research applications. This information has to be interpreted in a timely and accurate manner to benefit health care.

Medical image processing deals with the development of problem-specific approaches to the enhancement of raw medical image data for the purposes of selective visualization as well as further analysis. Medical image analysis then concentrates on the development of techniques to supplement the mostly qualitative and frequently subjective assessment of medical images by human experts with a variety of new information that is quantitative, objective, and reproducible.

What is perhaps most remarkable about these advances in medical imaging is the fact that the challenges have required significant innovation in computational techniques for nearly all aspects of image processing in various fields. The use of multiple imaging modalities on a single patient, for example MRI and PET, requires sophisticated algorithms for image registration and pattern matching.

Automated recognition and diagnosis require image segmentation, quantification and enhancement tools. Algorithms for image segmentation and visualization are employed broadly through many applications using all of the digital imaging modalities. The aim behind all these efforts is that by carefully tailoring and combining these image processing tools we create additional aids to the physicians in diagnosis, treatment planning, and disease monitoring and, thus, ultimately benefit patient care.

Standardization of the MRI Intensity Scale

A major difficulty with the MRI techniques for most protocols has been that image intensities do not have a fixed meaning, even within the same protocol, for the same body region, for images obtained on the same scanner, for the same patient. This poses problems in finding the proper window setting for image display, and more importantly, in image segmentation and analysis. Without the same protocol-specific intensity meaning, setting values for the parameters of image segmentation methods in a patient-study-independent fashion becomes very difficult. This results in compromising precision, accuracy, and efficiency of segmentation.

We recently developed an image processing method for MRI intensity standardization wherein all images (independent of patients and the specific brand of the MR scanner used) can be transformed in such a way that for the same protocol and body region, in the transformed images similar intensities will have similar tissue meaning. The method is described and extensively evaluated in Chapter 2 of the thesis and were published in [122, 125, 126, 127, 128, 130, 131, 135].

Image Segmentation

In brain MRI analysis, usually the first step is to segment the brain parenchyma and remove the extrameningeal tissues and bone. Also, in many situations, separating the gray matter and white matter is desirable, as well as separating normal and abnormal tissues such as normal appearing and abnormal normal appearing white matter, and lesions.

Image segmentation is the process of identifying and delineating objects in images. Segmentation is needed directly or indirectly for most of the operations done on images. It is also the most difficult of all image operations. Segmentation may be thought of as consisting of two related tasks—recognition (or detection) and delineation. In most recognition tasks, trained human operators outperform any computer algorithms. On the contrary, computer algorithms exist for delineation that are more precise, accurate, and efficient than human delineation of object regions or boundaries.

In automatic segmentation methods, a question arises as to what to do in case of failures. The premise in interactive methods is that, often a simple help from an operator is sufficient as a recognition aid. In clinical practice, the methods utilized in clinical trials invariably require some human assistance for each study for a human reader to check the results to insure that they are correct. Therefore, if we can make this process efficient, the uncertainties of the automatic methods can be overcome, and we have a solution that is practical.

The graded composition of tissues resulting from pathological heterogeneity and partial volume effects are not accurately accounted for by binary delineation methods. Fuzzy image processing frameworks become essential to handle material heterogeneity, noise, background intensity variation, fuzziness inherent in images, and how image elements hang together in image space and feature space in the presence of this fuzziness.

Image segmentation techniques using fuzzy connectedness principles have been used successfully in several large applications in recent years: MR brain image analysis for quantifying white matter lesions and brain tumors, segmentation of vessel structures and artery-vein separation in MRA, craniofacial muscle segmentation, and mammographic fibroglandular density quantification. However, there were several parts in these techniques where improvements can be achieved.

When the fuzzy connectedness algorithms were first introduced they could not be used for interactive processing due to their excessive computational requirements. In an attempt to substantially speed up the fuzzy connectedness methods we studied systematically a host of 18 'optimal' graph search algorithms. In Chapter 3 of the thesis, we demonstrate, by extensively testing these algorithms on a variety of 3D medical images taken from large on-going applications, that substantial improvement over current speeds is achievable with a combination of well-chosen algorithms and fast modern PCs. These results were published in [123, 124].

Despite the vast amount of literature and wide range of approaches to image segmentation, general, reusable frameworks seem to be lacking. General frameworks, that are readily applicable to images acquired with different protocols do not seem to have been published. Most papers deal with methods devised for specific tasks, tailored to a specific acquisition protocol and body region. In many cases, the segmentation method is a simple, *ad hoc* combination of basic tools or some more sophisticated combination involving heavy "engineering". What we need is a segmentation "workshop" wherein a protocol-specific segmentation method can be quickly fabricated. That is, for a new protocol, the workshop allows quick experimentation to complete the set up phase quickly rather than requiring months and years of basic research and development.

We devised one such workshop for brain MRI segmentation and present it in Chapter 4 of the thesis. The method (published in [132]) combines the robust, accurate, and efficient techniques of fuzzy connectedness segmentation with standardized MRI intensities [128, 135] (see Chapter 2 of the thesis) and fast algorithms [124] (see Chapter 3 of the thesis). The result is a general segmentation framework that efficiently utilizes the user input (for recognition) and the power of computer (for delineation). This same two-phase method has been applied to segment brain tissues from a variety of MRI protocols.

Image Registration

To be able to utilize the complementary information from multiprotocol acquisitions, or for longitudinal image analysis, the images need to be in spatial correspondence. Image registration is used to achieve this correspondence. Intra-subject image registration is used for matching multiple acquisitions for the same study or for matching longitudinal acquisitions. Inter-subject registration is used when creating atlases (of "average" anatomy), or when using the atlases in individual cases (e.g., for matching the atlas to the study in question and classifying the tissues based on the pre-determined classification of the atlas).

The various MR imaging methods provide somewhat disparate pieces of the information relating to the MS puzzle. The use of quantitative MRI in predicting the onset of MS and in assessing disease severity in established MS is actively pursued at present. In all such efforts, image registration to combine the complementary information from the images acquired under different protocols becomes vital. Many of the image segmentation methods for MS require one or more preprocessing operations, including registration to handle multichannel features. New quantitative measures are also being developed from the images for each of several tissue regions including the whole brain parenchyma, gray matter, white matter, lesions and cerebrospinal fluid (CSF), for patients with MS. These methods also depend on image registration.

In Chapter 5 of the thesis, we present a new class of approaches for rigid-body registration. We compare their performance with several commonly used strategies in studying Multiple Sclerosis via multi protocol MRI. The approach and the comparison results were published in [133, 134].

Image Analysis in Multiple Sclerosis

Multiple Sclerosis (MS) is an acquired disease of the central nervous system. Since its first description in 1877, it has been investigated extensively. In its advanced state, the disease may severely

impair the ambulatory ability and may even cause death. There are several clinical measures commonly used to express the severity of the disease. These measures are subjective and may have reproducibility problems. The development of new treatments demands objective outcome measures for relatively short trials. MR imaging has proven to be a very sensitive marker of the MS disease. A variety of protocols are being investigated for improving the conspicuity of the lesions in the images in the early stage and in the advanced stage for both the microscopic and macroscopic processes.

A review of MR image analysis techniques currently used in MS is given in Chapter 6 of the thesis. Also, the family of techniques using fuzzy connectedness for different protocols is described and evaluated. The content of this chapter was published in [129, 185].

Contributions of the Thesis

Contributions in the first group are described in Chapter 2 and were published in journal articles [128, 135] and proceedings papers [126, 127, 131].

1/1. We developed an image processing method (algorithms and theory for correctness) for MRI intensity standardization wherein all images can be transformed in such a way that for the same protocol and body region, in the transformed images similar intensities will have similar tissue meaning.

1/2. We showed through image display examples that the consistency of the brightness level and contrast of images is considerably improved after standardization and showed that the tissues have much more consistent intensity ranges if the new variants of the standardizing method are used. We demonstrated, in quantitative studies, that the scanner dependent intra- and inter-patient intensity variations are substantially reduced after standardization.

Contributions in the second group are described in Chapter 3 and were published in the journal article [123] and proceedings paper [124].

2/1. We studied systematically a host of 18 'optimal' graph search algorithms and demonstrated, by extensively testing these algorithms on a variety of 3D medical images taken from large on-going applications, that substantial improvement in speeding up the fuzzy connectedness methods is achievable with a combination of well-chosen algorithms and fast modern PCs.

2/2. We showed that utilizing efficient algorithms and careful selection of implementations can speed up the computation of fuzzy connectedness values by a factor of 16 to 29 (on the same hardware), as compared to the implementation previously used in applications utilizing fuzzy object segmentation.

2/3. We found that the label-setting algorithms are faster than the label-correcting algorithms in real applications, as the theoretical complexity analysis suggested. However, those using more complicated data structures are not the fastest in practice. We suggested a general guideline for selecting an algorithm depending on the structure of objects to be segmented and their intensity distribution.

Contribution in the third group is described in Chapter 4 and was published in the proceedings paper [132].

3. Using the fast algorithms proposed herein on fast hardware, interactive speed of fuzzy object segmentation is achievable. This leads to a new line of segmentation techniques. If images having intensities with tissue-specific meaning (such as standardized MR images) are utilized, most of the parameters for the segmentation method can be fixed once for all, all intermediate

data can be computed before the user interaction is needed and the user can be provided with more information at the time of interaction. The reliable recognition (assisted by human operators) and the accurate, efficient, and sophisticated delineation (automatically performed by the computer) can be effectively incorporated into a single interactive process.

Contributions in the fourth group are described in Chapter 5 and were published in the journal article [133] and proceedings paper [134].

4/1. We presented a new class of approaches, using local scale information, for rigid-body image registration. We compared the performance of the new scale-based registration methods with several commonly used strategies in studying Multiple Sclerosis via multi protocol MRI. We devised and used a new ranking system for the comparison of the registration methods.

4/2. We analyzed both intra- and inter-protocol registration results among the methods for 3D data of the head, acquired from 10 MS patients using 6 MRI protocols. Our analysis indicates that there is no "best" method, although there are significant differences in the performance of the methods, for the particular application of combining information from multiprotocol MR imagery for studying MS.

Contribution in the fifth group is described in Chapter 6 and was published in the book chapter [129], the journal article [185], and proceedings paper [184].

5. We reviewed MR image analysis techniques currently used in MS and we described and evaluated the family of techniques using fuzzy connectedness for different protocols.

Összefoglaló

Ez a disszertáció az orvosi képfeldolgozás különböző területeiből a képek szegmentálásával (3. és 4. fejezet), regisztrálásával (5. fejezet), valamint azok analízisével (6. fejezet) foglalkozik. A 2. fejezet egy olyan előfeldolgozó eljárást mutat be, amely a képek szegmentálásában és azok analízisében is lényeges szerepet kap. Bár dolgozatban a módszereket elsősorban egy élénken kutatott orvosi alkalmazás, a sclerosis multiplex (SM) vizsgálatának tükrében mutatjuk be alapvetően a mágneses magrezonancián alapuló felvételek (MRI) használatával, a bemutatott módszerek és eredmények egyéb betegségek tanulmányozása illetve más modalitások alkalmazása esetén is használhatók, sőt, nem orvosi alkalmazásokban is hasznosak lehetnek.

Képfeldolgozás

A képfeldolgozás, akár a robotika, a számítógépes látás, vagy akár az orvostudomány területén alkalmazva, azonos módon kezeli a képalkotási modalitások és alkalmazások bármelyikére jellemző elemeket, mint például a képalkotás geometriáját, lineáris transzformációkat, a frekvencia teret, digitális és folytonos tereket, a szegmentálást, hisztogram analízist és morfológiát. A képanalízis, függetlenül annak alkalmazási területétől, magába foglalja az előismeretek felhasználását, tulajdonságok osztályozását, modelleknek képekhez történő illesztését, az alakleírást és a mesterséges intelligencia számos általános problémját és módszerét. Mindezek ellenére, bár ezek a hagyományos megközelítések és általános alkalmazások fontosak, az orvosi képek és orvosi alkalmazások különleges természete különleges kezelésmódot igényel.

Orvosi képfeldolgozás

Az orvosi képek sajátos természete legalább annyira gyökerezik a képalkotás folyamatában, mint magukban a tárgyakban, amiről a felvételek készülnek. Az orvosi képalkotás technológiája forradalmian fejlődött az elmúlt három évtizedben. A viszonylag alacsony költségű számítási erőforrások, számítógépes technikák elterjedése, mint pl. a számítógépes tomográf (CT), a mágneses magrezonancián alapuló képalkotás (MRI), a digitális szubtrakciós angiográfia, az ultrahang valamint a nukleáris emisszión alapuló technikák (PET, SPECT) mind értékes kiegészítői a radiológus képalkotó eszköztárának, melyekkel a betegségek felismerése és diagnózisa egyre megbízhatóbbá vált. A két-, három- és többdimenziós adatok egyre több és részletesebb információt hordoznak a klinikai és kutatási alkalmazások számára egyaránt. Ahhoz, hogy ezek a klinikai gyakorlatban hasznosíthatók legyenek, ezt az információt pontosan és gyorsan kell kiértékelni.

Az orvosi képfeldolgozás a képadat javítására, megjelenítésére és analízisére irányuló problémaspecifikus módszerek kidolgozásával foglalkozik. Az orvosi képfeldolgozás olyan technikák kifejlesztésére összpontosít, amelyek az orvosi képeknek a szakértők általi többnyire kvalitatív és gyakran szubjektív vizsgálatait új, kvantitatív, objektív és reprodukálható információkkal egészítik ki.

Ezen fejlesztésekkel kapcsolatban talán az a legfigyelemreméltóbb, hogy a kihívások lényeges újításokat követeltek a számítási technikák terén a képfeldolgozás majd minden ágán. Több képalkotó modalitás alkalmazásakor egyazon beteg esetén (például MRI és PET) a képek regisztrálására

és az alakfelismerés bonyolult algoritmusaira van szükség. Az automatikus felismeréshez és diagnózishoz elengedhetetlen a képek szegmentálása, a kvantifikálás és a képjavítás. A képek szegmentálására és megjelenítésére kidolgozott algoritmusok valamennyi digitális képalkotó modalitás esetében alkalmazásra kerülnek. Mindezen erőfeszítések végső célja az, hogy e képfeldolgozó eszközöket körültekintően hangolva, testreszabva és kombinálva az orvosok számára segédletet készítsünk a diagnózishoz, a kezelés tervezéshez és a betegség folyamatos ellenőrzéséhez és ezáltal a betegek gyógyítását segítsük.

Az MRI intenzitás skála standardizálása

A mágneses magrezonancián alapuló képalkotó módszer (MRI) gyökeresen változást hozott a radiológiában. Az MRI az emberi test belső felépítéséről noninvazív módon képes jó minőségű képeket készíteni és nincsenek jelenleg kimutatható egészségre káros mellékhatásai. Számos MRI protokollt dolgoztak ki mind kontrasztanyagok használatával és anélkül. Ezen protokollok használatával ugyanazon szervrendszerben különböző szöveteket különböző kontrasztviszonyokkal képezhetünk le. Sajnos ezeknél a technikáknál a felvételi protokollok többségénél felmerül az a probléma, hogy a képpontok intenzitásértékei nem rendelkeznek konkrét jelentéssel (ellentétben pl. a számítógépes tomográfiával (CT)), még akkor sem, ha ugyanazon a szkenneren ugyanazon protokoll szerint készítünk felvételeket ugyanazon beteg ugyanazon testrészéről.

Ez komoly problémákat jelent a képek megjelenítésekor, de ami még lényegesebb, azok szegmentálásában és kvantitatív analízisében is. A képfeldolgozási módszerek szinte kivétel nélkül paraméterek megfelelő beállítását igénylik. Ezek beállítása lényegesen nehezebb, ha a képpontok értékei nem bírnak jól meghatározott protokoll- illetve szövet-specifikus jelentéssel. Ez a szegmentálás reprodukálhatóságát, pontosságát és hatékonyságát veszélyezteti.

Kidolgoztunk egy olyan képfeldolgozó eljárást az MRI intenzitás skálájának standardizálására, amellyel a képek (a betegtől és az alkalmazott MR szkenner konkrét típusától függetlenül) egyszerűen transzformálhatók úgy, hogy az eredmény képek lényegesen hasonlóbb intenzitás jelentéssel bírnak, mint az eredeti képek. A módszert, melyet a [122, 125, 126, 127, 128, 130, 131, 135] dolgozatokban publikáltunk, a disszertáció 2. fejezetében ismertetjük és vizsgáljuk.

Képek szegmentálása

Agy MRI elemzésekor általában az első lépés az agy szegmentálása és az agyhártyán kívüli szövetek valamint a csontok eltávolítása. Gyakran szükség van a szürke- és fehérállományok elkülönítésére is, valamint a normális és abnormális szövetek, mint pl. az egészséges fehérállomány és laesiok közötti megkülönböztetésre.

A szegmentálás célja, hogy képekből az azokon leképezett objektumokról struktúrális információt nyerjünk ki. Szegmentálásra, közvetlenül vagy közvetve, a képfeldolgozó műveletek többségénél szükség van. A szegmentálás egyben a képeken végzett műveletek közül a legnehezebb is. A szegmentálást tekinthetjük két egymással összefüggő feladat, a felismerés (detection) és a körülhatárolás (delineation) együttesének. A felismerés a magasabb szintű feladat, amelyben az objektum hozzávetőleges elhelyezkedését határozzuk meg a képen. A körülhatárolás az alacsonyabb szintű feladat, amelyben az objektum pontos kiterjedését és pontonkénti összetételét határozzuk meg. A legtöbb felismerési feladatban egy képzett szakember lényegesen jobban teljesít, mint a számítógépes algoritmusok. Ezzel szemben léteznek olyan számítógépes algoritmusok, amelyek az embernél precízebben, pontosabban és hatékonyabban elvégzik az objektumok illetve azok határainak kijelölését.

Automatikus szegmentálási módszerek esetén felmerül a kérdés, hogy mit tegyünk a hibák esetén. Az interaktív módszerek előnye az, hogy gyakran az operátor egy egyszerű segítsége elegendő arra, hogy lényegesen előremozdítsa a felismerési folyamatot. A klinikai gyakorlatban használt módszerek kivétel nélkül valamilyen emberi közreműködésre épülnek. Az operátor ellenőrzi az eredményeket,

így biztosítva azok helyességét. Ezért, ha ezt a folyamatot képesek vagyunk hatékonyabbá tenni, legyőzhetjük az automatikus módszerek bizonytalanságait és egy praktikus megoldáshoz jutunk.

A szövetek változékony intenzitás összetételét, mely patológiai heterogenitásból illetve a részleges térfogat átlagolás jelenségből származik, a bináris körülhatárolási módszerek nem kezelik le pontosan. A fuzzy képfeldolgozó rendszerek alapvető eszközök az anyag heterogenitása valamint a zaj és háttér intenzitás változások lekezeléséhez.

Az elmúlt években a fuzzy összefüggőség elméletén alapuló szegmentálási technikákat számos komoly feladatban alkalmazták. A módszer azonban több ponton is javításra érett. Amikor a fuzzy összefüggőség algoritmusait először bevezették, azok a nagy számításigényük miatt nem voltak alkalmasak arra, hogy interaktív eljárásokban alkalmazzák őket.

A disszertáció 3. fejezetében összehasonlítunk 18 különböző gráfkeresési algoritmust fuzzy objektumok számítására 3D orvosi képekből, melyeket hosszútávú vizsgálatsorozatokból vettünk. Megmutatjuk, hogy az algoritmusok körültekintő megválasztásával az orvosi rutin feladatokban alkalmazott szegmentálások lényegesen felgyorsíthatók. Ha az új, hatékony algoritmusokat alkalmazzuk modern, gyors PC-ken, interaktív szegmentálás is elérhető a fuzzy összefüggőség módszerét alkalmazva. Ezeket az eredményeket a [123, 124] dolgozatokban publikáltuk.

Annak ellenére, hogy szegmentálási módszerek széles sora található a szakirodalomban, úgy tűnik, hiányoznak az általános, újrafelhasználható rendszerek, melyeket különböző protokollokkal felvett képekre készen lehetne alkalmazni. A legtöbb közlemény egy-egy konkrét, protokoll- és testrész-specifikus feladatra készült módszert ismertet. A szegmentálási eljárás gyakran egyszerű eszközöknek az adott feladatra történő összebarkácsolásából alakul ki. Amire igazán szükség lenne, az egy olyan szegmentálási keretrendszer, amelyben a protokoll-specifikus módszerek gyorsan kifejleszthetők lennének. A keretrendszer használatával gyors kísérletekre lenne lehetőség az új protokollokhoz, és igy a konkrét módszer előkészítését gyorsan elvégezhetnénk a több hónapos (esetleg éves) alapkutatások és fejlesztések helyett.

Egy ilyen rendszert dolgoztunk ki az agy szegmentálására MRI képeken, melyet a disszertáció 4. fejezetében mutatunk be és a [132] dolgozatban publikáltunk. A módszer a robusztus, pontos és hatékony fuzzy összefüggőségen alapuló szegmentálási technikát ötvözi standardizált MRI intenzitás skálával (lásd a disszertáció 2. fejezete és [128, 135]) és gyors algoritmusokkal (lásd a disszertáció 3. fejezete és [124]). Eredményként egy olyan általános szegmentáló rendszert kaptunk, amely hatékonyan ötvözi a felhasználói segítséget (a felismerési műveletben) a számítógép erőforrásaival (a körvonalazáshoz). Ugyanerre a kétfázisú módszerre építve kidolgoztunk egy eljárást az agy szöveteinek szegmentálására sclerosis multiplex (SM) betegek felvételeiből több, a napi rutinban alkalmazott MR protokoll esetére.

Képek regisztrációja

A többprotokollos felvételekből származó egymást kiegészítő információk együttes alkalmazásához illetve a képeken alapuló hosszútávú vizsgálatokhoz szükség van arra, hogy a képek térben megfeleltethetők legyenek egymásnak. A képregisztráció feladata, hogy több, ugyanazon objektumról készített felvételt egy közös koordinátarendszerben ábrázoljunk. Objektumon belüli képregisztrációt alkalmazunk akkor, amikor ugyanazon vizsgálat különböző felvételeit, vagy amikor ugyanazon szerv időben elérő felvételeit illesztjük. Objektumok közötti képregisztrációt használunk anatómiai atlaszok előállításakor illetve amikor az atlaszokat konkrét egyedi esetekre alkalmazzuk.

A változatos MRI képalkotási módszerek egymástól különböző, egymást kiegészítő információkat szolgáltatnak az SM „mozaikhoz”. A kvantitatív MRI használata az SM kialakulásának megjósolásában, illetve a betegség előrehaladottságának megállapításában az utóbbi években aktívan kutatott terület. Minden ilyen irányú törekvésben létfontosságú feladat a képek regisztrálása, ugyanis csak így lehet összekapcsolni a különböző protokollokkal felvett képekből az egymást kiegészítő információkat. Az SM vizsgálatában alkalmazott képszegmentáló módszerek többségéhez szükség van egy vagy több előfeldolgozó műveletre, köztük a regisztrációra is a többcsatornás jellemzők

kezeléséhez. Az SM felvételekből új kvantitatív mértékek kifejlesztését célzó eljárások szintén építenek a képek regisztrációjára.

A disszertáció 5. fejezetében a merevtest regisztrációk egy új osztályát mutatjuk be és hasonlítjuk össze hagyományos, széles körben alkalmazott módszerekkel egy konkrét alkalmazási környezetben, a sclerosis multiplex többprotokollos MRI-vel történő vizsgálatában. A módszert és az összehasonlítás eredményeit a [133, 134] dolgozatokban publikáltuk.

Képanalízis a sclerosis multiplex vizsgálatában

A sclerosis multiplex (SM) a központi idegrendszer idült, gyulladásos betegsége, mely annak 1877-es első leírása óta élénken kutatott terület. A betegség előrehaladott állapotban komolyan károsíthatja a beteg mozgásképességét és akár halált is okozhat. A jelenleg használt klinikai mértékek, melyekkel a betegség fokozatait mérik, sajnos szubjektív és megismételhetőségi problémákkal bírnak. Új kezelések kifejlesztéséhez objektív mértékekre van szükség még viszonylag rövid idejű vizsgálatokhoz is. Az MRI az SM diagnosztizálásának és a betegség nyomonkövetésének egyik legfontosabb klinikai tesztje. Számos MRI protokollt vizsgálnak azzal a céllal, hogy a laesiok mind a korai mind az előrehaladott szakaszban jobban láthatók legyenek a képeken.

A disszertáció 6. fejezetében áttekintjük az SM tanulmányozásában jelenleg használatos MR képanalízis technikákat valamint bemutatjuk és elemezzük a különböző protokollokra kidolgozott, fuzzy összefüggőségen alapuló technikákat. Az itt leírtakat a [129, 185] dolgozatokban publikáltuk.

A disszertáció eredményei

Az első téziscsoport eredményeit a disszertáció 2. fejezetében mutatjuk be és a [128, 135] folyóiratcikkekben valamint a [126, 127, 131] proceedings közleményekben publikáltuk.

1/1. Kidolgoztunk egy olyan képfeldolgozó eljárást az MRI intenzitás skálájának standardizálására, amellyel a képek (a betegtől és az alkalmazott MR szkenner konkrét típusától függetlenül) egyszerűen transzformálhatók úgy, hogy az eredmény képek lényegesen hasonlóbb intenzitás jelentéssel bírnak, mint az eredeti képek.

1/2. Példákon keresztül megmutattuk, hogy a képek fényességi és kontraszt tulajdonságai lényegesen javulnak a standardizálás után valamint azt, hogy a szövetek sokkal konzisztensebb intenzitás tartományokkal bírnak, amennyiben az standardizálási módszer újabb változatait alkalmazzuk. Kvantitatív elemzésekkel megmutattuk, hogy az azonos beteg képei illetve különböző betegek képei közötti a szkenner-függő intenzitás változékonyság lényegesen csökken standardizálás után.

A második téziscsoport eredményeit a disszertáció 3. fejezetében mutatjuk be és a [123] folyóiratcikkben valamint a [124] proceedings közleményben publikáltuk.

2/1. Összehasonlítottunk 18 különböző gráfkeresési algoritmust fuzzy objektumok számítására 3D orvosi képekből. Megmutattuk, hogy az algoritmusok megfelelő megválasztásával és hatékony PC-k alkalmazásával az orvosi rutin feladatokban alkalmazott szegmentálások lényegesen felgyorsíthatók.

2/2. Megmutattuk, hogy hatékony algoritmusok alkalmazásával és azok megvalósításának körültekintő megválasztásával a fuzzy összefüggőségi értékek számítása a korábbi alkalmazásokban használt változatokénak 16–29-szeresére felgyorsítható ugyanazon a hardveren.

2/3. Tapasztalataink szerint a cimke-beállító algoritmusok valós alkalmazásokban gyorsabbak mint a cimke-javító algoritmusok, ahogy azt az elméleti elemzésük is jelezte. Azok az algoritmusok viszont, amelyek bonyolultabb adatstruktúrákat alkalmaznak, nem mindig gyorsabbak a

gyakorlatban. Megadtunk egy általános útmutatót arra, hogy a szegmentálandó objektumok struktúráját és azok intenzitásainak eloszlását figyelembe véve konkrét alkalmazások esetén hogyan válasszuk ki a megfelelő algoritmust.

A harmadik téziscsoport eredményeit a disszertáció 4. fejezetében mutatjuk be és a [132] proceedings közleményben publikáltuk.

3. Az általunk javasolt gyors algoritmusokat gyors hardveren alkalmazva a fuzzy objektumok interaktív sebességel szegmentálhatók. Ez a szegmentálási technikák egy új vonalát nyitja meg. Ha olyan képeket használunk, amelyekben az intenzitások szövet-specifikus jelentéssel bírnak (mint pl. a standardizált MR képek), a szegmentálási eljárás paraméreinek többsége előre rögzíthető, a köztes adatok előre kiszámíthatók és az interakciós fázisban a felhasználó számára több információ áll rendelkezésre. A megbízható, szakember által támogatott felismerés és a pontos, hatékony, összetett, automatikus számítógépes körülhatárolás egyetlen interaktív eljárásba foglalható.

A negyedik téziscsoport eredményeit a disszertáció 5. fejezetében mutatjuk be és a [133] folyóiratcikkben valamint a [134] proceedings közleményben publikáltuk.

4/1. Bemutattunk egy új megközelítést merev-test regisztrációra, amely a lokális skála információra épül. Összehasonlítottuk az új skála-alapú regisztrációs módszert néhány széles körben alkalmazott módszerrel egy konkrét alkalmazás tükrében, amely az SM vizsgálata többprotokollos MRI-vel. Kidolgoztunk és alkalmaztunk egy új sémát a regisztrációs módszerek összehasonlítására.

4/2. Mind a protokollon belüli mind a protokollok közötti regisztrációs eredményeket 10 SM beteg agyáról 6 MRI protokollal készült 3D képadaton vizsgáltuk. Eredményeink azt mutatják, hogy nincs egyértelműen legjobb módszer, bár vannak szignifikáns eltérések a módszerek működése között a fent említett alkalmazás esetén.

Az ötödik téziscsoport eredményeit a disszertáció 6. fejezetében mutatjuk be és a [129] könyv fejezetben, a [185] folyóiratcikkben valamint a [184] proceedings közleményben publikáltuk.

5. Áttekintettük az SM vizsgálatára jelenleg használatos MR képanalízis technikákat és bemutattuk és kiértékeltük a fuzzy összefüggőségen alapuló módszerek családját különböző protokollokra.

Publications

Chapters in books

L. G. Nyúl and J. K. Udupa. MR image analysis in multiple sclerosis. In J. A. Frank, editor, *Advances in Multiple Sclerosis*, volume 10(4) of *Neuroimaging Clinics of North America*, pages 799–816. W. B. Saunders Company, Philadelphia, PA, November 2000.

Articles in peer reviewed journals

I. Catalaa, R. I. Grossman, D. L. Kolson, J. K. Udupa, L. G. Nyúl, L. Wei, X. Zhang, M. Polansky, L. J. Mannon, and J. C. McGowan. Multiple sclerosis: Magnetization transfer histogram analysis of segmented normal-appearing white matter. *Radiology*, 216(2):351–355, August 2000.

Y. Ge, R. I. Grossman, J. K. Udupa, J. S. Babb, L. G. Nyúl, and D. L. Kolson. Brain atrophy in relapsing-remitting multiple sclerosis: Fractional volumetric analysis of gray matter and white matter. *Radiology*, 220(3):606–610, September 2001.

Y. Ge, J. K. Udupa, L. G. Nyúl, L. Wei, and R. I. Grossman. Numerical tissue characterization in MS via standardization of the MR image intensity scale. *Journal of Magnetic Resonance Imaging*, 12(5): 715–721, November 2000.

L. G. Nyúl, A. X. Falcão, and J. K. Udupa. Fuzzy-connected 3D object segmentation at interactive speeds. *Graphical Models*, 1999. submitted for publication.

L. G. Nyúl and J. K. Udupa. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine*, 42(6):1072–1081, December 1999.

L. G. Nyúl, J. K. Udupa, and P. K. Saha. Task-specific comparison of 3-D image registration methods. *IEEE Transactions on Medical Imaging*, 2001. submitted for publication.

L. G. Nyúl, J. K. Udupa, and X. Zhang. New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–150, February 2000.

J. K. Udupa, L. G. Nyúl, Y. Ge, and R. I. Grossman. Multiprotocol MR image segmentation in multiple sclerosis: Experience with over 1000 studies. *Academic Radiology*, 8(11):1116–1126, November 2001.

Full papers in conference proceedings

L. G. Nyúl, A. X. Falcão, and J. K. Udupa. Fuzzy-connected 3D image segmentation at interactive speeds. In K. M. Hanson, editor, *Medical Imaging 2000: Image Processing*, volume 3979 of *Proceedings of SPIE*, pages 212–223, June 2000.

L. G. Nyúl and J. K. Udupa. An approach to standardizing the MR image intensity scale. In S. K. Mun and Y. Kim, editors, *Medical Imaging 1999: Image Display*, volume 3658 of *Proceedings of SPIE*, pages 595–603, May 1999.

L. G. Nyúl and J. K. Udupa. New variants of a method of MRI scale normalization. In A. Kuba, M. Šámal, and A. Todd-Pokropek, editors, *Information Processing in Medical Imaging: 16th International Conference, IPMI'99, Proceedings*, volume 1613 of *Lecture Notes on Computer Science*, pages 490–495, Berlin, June–July 1999.

L. G. Nyúl and J. K. Udupa. Standardizing the MR image intensity scales: Making MR intensities have tissue-specific meaning. In S. K. Mun, editor, *Medical Imaging 2000: Image Display*, volume 3976 of *Proceedings of SPIE*, pages 496–504, April 2000.

L. G. Nyúl and J. K. Udupa. A protocol-independent brain MRI segmentation method. In M. Sonka and J. M. Fitzpatrick, editors, *Medical Imaging 2002: Image Processing*, volume 4684 of *Proceedings of SPIE*, pages 1588–1599, May 2002.

L. G. Nyúl, J. K. Udupa, and P. K. Saha. Task-specific comparison of 3-D image registration methods. In M. Sonka and K. M. Hanson, editors, *Medical Imaging 2001: Image Processing*, volume 4322 of *Proceedings of SPIE*, pages 1588–1598, July 2001.

J. K. Udupa, L. G. Nyúl, Y. Ge, and R. I. Grossman. Multiprotocol MR image segmentation in multiple sclerosis: Experience with over 1000 studies. In K. M. Hanson, editor, *Medical Imaging 2000: Image Processing*, volume 3979 of *Proceedings of SPIE*, pages 1017–1027, June 2000.

Conference abstracts and extended abstracts

I. Catalaa, R. I. Grossman, D. L. Kolson, L. G. Nyúl, L. Wei, J. K. Udupa, M. Polansky, and J. C. McGowan. Magnetization transfer histogram analysis of segmented normal-appearing white matter in multiple sclerosis. In *ISMRM Seventh Meeting Proceedings, 22-28 May 1999, Philadelphia, PA*, page 957, May 1999.

Y. Ge, J. K. Udupa, L. G. Nyúl, L. Wei, and R. I. Grossman. Numerical tissue characterization in MS via standardization of the MR image intensity scale. In *ISMRM Eighth Meeting Proceedings, 1-7 April 2000, Denver, CO*, volume 1, page 579, April 2000.

Y. Ge, R. I. Grossman, J. K. Udupa, J. S. Babb, L. G. Nyúl, and J. C. McGowan. Tissue characterization in relapsing-remitting and secondary-progressive MS via magnetization transfer ratio. In *ISMRM Eighth Meeting Proceedings, 1-7 April 2000, Denver, CO*, volume 2, page 1189, April 2000.

L. Nyúl and J. K. Udupa. Többdimenziós MRI képek feldolgozása. In A. Kuba, E. Máté, and K. Palágyi, editors, *Book of Abstracts, KEPAF 2002, Domaszék, Hungary, 23-25 January, 2002*, pages 96–97, January 2002.

L. G. Nyúl and J. K. Udupa. On standardizing the MR image intensity scale. *Radiology*, 209(P): 581–582, November 1998.

L. G. Nyúl and J. K. Udupa. Standardizing the MR image intensity scale and its applications. In T. Csendes, editor, *Book of Abstracts, Conference of PhD Students in Computer Science, July 20-23, 2000, Szeged, Hungary*, page 75, July 2000.