

Clustering of fiber tracts and quantitative analysis of white matter fiber bundles in the human brain

DISSERTATION

Zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

vorgelegt der Fakultät für Informatik und Automatisierung
der Technischen Universität Ilmenau

von

Dipl.-Inf. Christian Ros

geboren am 8. April 1980 in Suhl

Tag der Einreichung: 13.01.2014

Tag der wissenschaftlichen Aussprache: 14.01.2015

1. Gutachter: Prof. Dr.-Ing. habil. Jens Haueisen
2. Gutachter: Prof. Dr. rer. nat. med. habil. Jürgen R. Reichenbach
3. Gutachter: PD Dr.-Ing. habil. Thomas R. Knösche

Acknowledgment

This thesis would have not been possible without the help and support of great people to whom I dedicate this chapter.

Foremost I thank my supervisors and mentors Prof. Jens Haueisen and Prof. Jürgen R. Reichenbach for giving me the opportunity to pursue a doctorate and contribute to science. I am especially grateful to Prof. Reichenbach who has been a great advisor over the years – not only on a scientific but also on a personal level. Your continuous support, encouragement, trust and excellent guidance but also your constructive criticism were priceless and helped me all along the way. Thank you for teaching me so much about science and for giving me the freedom to explore and pursue my own ideas.

I would also like thank my colleagues and collaborators from the Pediatric Radiology, Prof. Hans-Joachim Mentzel and Dr. Martin Stenzel who supported me throughout the years and introduced me to the fascinating realm of pediatric radiology. Thank you for our rich and invaluable discussions that greatly helped to improve my radiological knowledge. On top of that, I also thank all my colleagues from the Center of Radiology. Special thanks go to my former co-worker Dr. Hartmut Peter Burmeister who often made my day by recounting great stories of our previously experienced events at the institute. I thank Ines Krumbein for giving me valuable insight on human brain anatomy and for pinpointing all sorts of brain structures with unmatched accuracy. I am also grateful to all our collaborators at the Institute of Psychology of the Friedrich-Schiller-University Jena, at the Department of Neurology and at the Department of Psychiatry and Psychotherapy from the Jena University Hospital, who contributed to our joint project, acquired the data and provided exceptional expertise that helped to focus our research on the important tasks at hand.

I thank my colleague Dr. Karl-Heinz Herrmann with whom I shared the office. I am grateful for your continuous support and for our enlightening conversations about all sorts of topics which broadened my scientific horizon greatly. Thank you for introducing me to the mysteries and pitfalls of the *Linux operating system*, for showing me how to prepare excellent espresso and for memorable road trips on Maui and in Australia after visiting the ISMRM. I also thank Dr. Daniel Güllmar for supporting me during the doctorate and for our fruitful discussions that always provided new ideas and impulses. The IT infrastructure that you have carefully assembled and improved over the years is a rare treasure that helped me tremendously to process and analyze a huge amount of data.

Special thanks to my fellow doctoral students and those who have graduated during my time in the Medical Physics Group: Dr. Alexander Gussew, Kevin Tschiesche, Marianne Cleve,

Patrick Hiepe, Martin Krämer, Dr. Ferdinand Schweser, Dr. Andreas Deistung and Xiang Feng. I am grateful for our stimulating and inspiring discussions, for sharing ideas and enthusiasm. Our coffee breaks brought us all together and were not only invaluable to develop and frame new ideas but also to uncover and dismiss our follies.

It is a personal concern of mine to thank all recent and past members of the Medical Physics Groups for wonderful times and for providing such a pleasant environment to work in. Over the years, the Medical Physics Group has become home and with all your quirks, you guys (and girls) have become family. I certainly will always be delighted to remember our great times with BBQs, beer, cake, *Mettigel* and “in-room” badminton. While we participated in conferences and visited amazing cities all around the world, we gathered not only a great amount of knowledge but also experienced remarkable adventures. I am very grateful for all of this and especially to have been part of the Medical Physics Group.

I am most grateful to my family and all my friends who encouraged me to pursue my goals and supported me with all their strength. To my parents Sylvia and Bernd Ripperger I am particularly indebted. Your unlimited love and dedicated support in every way imaginable made this research possible and made me who I am. I thank my grandfather Alfred for always being proud of me and for always believing in me. As he unfortunately passed away when I started my research, I dedicate this thesis to his memory. May he rest in peace.

My deepest and most sincere thanks go to Maggi, my beloved girlfriend, who endured this doctorate with me. Without your endless support and ceaseless encouragement, writing this thesis would have been an unimaginably difficult endeavor. I am grateful to you for believing in us, for supporting me over the years and for your extraordinary patience. Thank you for being there for me and for enriching my life in unthinkable ways!

— Christian Ros
Jena, on an exceptionally warm (and late) evening in January 2014

Zusammenfassung

Das Gehirn ist das neuronale Zentrum des menschlichen Körpers und besitzt eine höchst komplexe Mikrostruktur aus Nervengewebe mit untereinander verknüpften Nervenzellen. Die Nervenzellen sind, je nach ihren Aufgaben (wie visuelle Wahrnehmung, motorische Fähigkeiten oder Sprache), in funktionellen Einheiten angeordnet und über ihre Axone miteinander verbunden. Die Axone bilden dabei ein weit verzweigtes neuronales Netzwerk, das Nervenzellen funktionell gleicher Areale untereinander verbindet und die Signalübertragung zwischen den verschiedenen funktionellen Einheiten auch über lange Strecken sicherstellt.

Die Entwicklung der *Magnetresonanztomographie* und insbesondere die Entwicklung der *diffusionsgewichteten Bildgebung* ermöglichten es, das neuronale Netzwerk des menschlichen Gehirns in bisher ungeahnter Weise nicht-invasiv zu studieren um dessen immanente Komplexität zu entschlüsseln. Durch den Einsatz von Traktographieverfahren können drei-dimensionale Fasertrakte rekonstruiert werden, welche den Verlauf der zugrunde liegenden, mikroskopischen Nervenfasern modellieren. Das daraus resultierende, komplexe Netzwerk von Faserverbindungen approximiert die strukturelle Organisation des neuronalen Netzwerks und erlaubt Rückschlüsse über den Verlauf und die Konnektivität der funktionellen, untereinander verbundenen Areale zu ziehen.

Aufgrund der Komplexität des extrahierten Fasernetzwerks und der Ungeordnetheit der darin enthaltenen Fasertrakte ist eine zeitaufwendige und vielschichtige Prozessierung der gewonnenen Daten notwendig, was die Einsetzbarkeit der Traktographie für viele medizinische Anwendungen außerordentlich einschränkt. Um die Anwendungsmöglichkeiten der Traktographie zu verbessern, werden im Rahmen dieser Dissertation neue Konzepte und Strategien vorgestellt, die es ermöglichen Fasertrakte mit Hilfe der Clusteranalyse vollautomatisch zu Faserbündeln zusammenzufassen, welche dem Verlauf der zugrunde liegenden mikrostrukturellen Faserbahnen entsprechen. Hierfür wird mit CATSER (*cluster analysis through smartly extracted representatives*) eine neue Methode für die automatische Clusteranalyse der Fasertrakte vorgestellt. Diese nutzt die intrinsische Redundanz der Daten, um auch die Analyse großer Datensätze zu ermöglichen. Um die Korrespondenz zwischen den resultierenden Faserbündeln und den zugrunde liegenden mikrostrukturellen Faserbahnen weiter zu verbessern, können zusätzliche anatomische Informationen eines Atlanten der weißen Substanz von CATSER berücksichtigt werden. Durch die Nutzung paralleler Rechnerarchitekturen, neuer Algorithmen und Ähnlichkeitsmaße kann die Clusteranalyse in einem vertretbaren Zeitrahmen realisiert werden. Verschiedene Experimente wurden durchgeführt, um die Eigenschaften des vorgestellten Clusteranalyseverfahrens zu untersuchen und seine hohe Performanz nachzuweisen.

Die schnelle und zuverlässige Extraktion der Faserbündel mit CATSER eröffnet einer Vielzahl medizinischer Anwendungen die Nutzung traktographischer Daten. Gruppen-

basierte Analysemethoden die verwendet werden, um veränderte Diffusion zwischen verschiedenen Gruppen (wie gesunden Probanden vs. Patienten) oder zwischen Untergruppen von Patienten zu untersuchen, sind dabei prädestiniert, um von den zusätzlichen anatomischen Informationen der Faserbündel zu profitieren. Aus diesem Grund präsentiert der zweite Teil der Dissertation ein neues Verfahren für die quantitative Analyse von Diffusionsparameter, welches die Analyse durch Einbeziehung der vorab extrahierten Faserbündel verbessert. Diese neue Technik ermöglicht die voxelbasierte Untersuchung individueller Bündel, beschränkt die Analyse dabei allerdings ausschließlich auf Voxel, die zu dem untersuchten Faserbündel gehören. Überlappende Bereiche anderer Bündel, die nicht Gegenstand der Untersuchung sind, werden mit dieser Technik effektiv ausgeblendet. Um die Anwendbarkeit der vorgestellten Methoden zu eruieren, wurden die neuen Techniken im Rahmen von zwei Studien eingesetzt. Bei gesunden Probanden und schizophrenen Patienten wurde die Diffusion in ausgewählten Faserbündeln der linken und rechten Hemisphäre untersucht und verschiedene Diffusionsparameter miteinander verglichen. Hierbei konnten beide Studien Unterschiede in bestimmten Faserbündeln der linken und rechten Hemisphäre nachweisen und dabei das Potential der vorgestellten Techniken erfolgreich demonstrieren.

Abstract

The human brain as the neural processing center consists of a highly complex microstructure that is primarily composed of interconnected nerve cells. While the nerve cells are arranged in functional units, dedicated to perform certain unique tasks (e.g., visual perception, motor control or speech), their axons form a complex network that interconnects nerve cells of the same area and enables long-range signal propagation between different functional units.


With the advent of *magnetic resonance imaging* and, in particular, the introduction of *diffusion weighted imaging* researchers began to study the fiber network of the human brain, pursuing the goal to unravel its inherent complexity non-invasively. The employment of tractography facilitated this process and enabled the approximation of the structural organization of the fiber network with three-dimensional fiber tracts, which opened the unique opportunity to study the course and the integrity of the microstructure in unprecedented ways.

For many medical applications, the automatic parcellation of fiber tracts into bundles that represent the underlying microstructure is of extraordinary importance, but often too time consuming and too complicated, thus limiting the usability of fiber tractography. In order to improve the applicability of fiber tractography, this thesis introduces novel concepts and strategies towards automatic parcellation of fiber tracts by using cluster analysis. With CATSER (*cluster analysis through smartly extracted representatives*), a new method for the automated clustering of fiber tracts is presented that exploits the intrinsic redundancy of the data to make cluster analysis applicable to large tractography datasets. In order to improve the grouping of tracts into bundles that represent the underlying microstructure more correctly, CATSER can be used together with a white matter atlas. By using parallel computing and random sampling as well as novel similarity measures, the cluster analysis of fiber tracts can be performed in reasonable time. To study properties of the introduced cluster analysis framework and to demonstrate its performance in a multiprocessing environment, various experiments were conducted.

The fast and consistent extraction of fiber bundles with CATSER aids a multitude of medical applications. Group-based analyses methods that are used to study altered diffusion between groups of subjects (e.g., healthy volunteers vs. patients) or between subgroups of patients are particularly suited candidates to benefit from the additional anatomical information of fiber bundles. For this reason, this thesis introduces a novel *fiber bundle-driven technique* that takes advantage of fiber bundles to enhance quantitative group-based analyses. This new technique enables selective, voxel-wise analysis of local diffusivity properties for individual bundles and restricts the analysis to voxels that belong to the *bundle of interest*. Contributions of other regions or bundles that are not part of the analyzed bundle are thereby suppressed. To investigate the applicability of these new methods, diffusion-associated hemispheric differences were studied in selected fiber bundles of healthy volunteers and schizophrenic patients.

Alterations of diffusion-related parameters between corresponding bundles of the left and right hemisphere were detected in various bundles of both groups. With these two studies the potential of the novel techniques was successfully demonstrated.

Acronyms and abbreviations

$\text{abs}(\mathbf{x})$	absolute value of \mathbf{x}
α	flip angle
$\det(\mathbf{D})$	determinant of \mathbf{D}
\wedge	binary and
\mathbb{D}	dataset that consists of $ \mathbb{D} $ elements
\mathbf{G}	ground truth dataset
\mathbf{S}	$N \times N$ similarity matrix for a dataset \mathbb{D} with $N = \mathbb{D} $ elements
$\ \mathbf{x}\ $	2-norm of \mathbf{x}
\oplus	binary xor (either or)
$\rho(\mathbf{v})$	tract density in a voxel \mathbf{v}
$ X $	cardinality of a set X
$\tau(\mathbf{x})$	quantitative value in \mathbf{x} , where \mathbf{x} is a voxel or a point in 3D space
\mathbf{D}	symmetric, second order diffusion tensor
ω_o	Larmor frequency
\mathbf{B}_o	main magnetic field
\mathbf{g}	gradient \mathbf{g}
$\zeta(A, C)$	spatial resemblance between atlas class A and cluster C
g	gradient strength of a gradient \mathbf{g}
P_F	feature processing order
P_{SM}	similarity measure processing order
T_1	T_1 relaxation time (spin-lattice relaxation time)
T_2	T_2 relaxation time (spin-spin relaxation time)
T_E	echo time
T_R	repetition time
2D	two-dimensional
3D	three-dimensional
	brain
ADC	apparent diffusion coefficient
ANTs	advanced normalization tools
AP	anterior-posterior
ATLAS	automatically tuned linear algebra software
ATR	anterior thalamic radiation
BLAS	basic linear algebra subprograms
BOI	bundle of interest
CATSER	cluster analysis through smartly extracted representatives
CD	combined distance

Acronyms and abbreviations

CGC	gyrus part of the cingulum
CGH	hippocampal part of the cingulum
CNS	central nervous system
CPU	central processing unit
CST	corticospinal tract
CUDA	compute unified device architecture
CURE	clustering using representatives
DS	dataset
DTI	diffusion tensor imaging
DTK	diffusion toolkit
DWI	diffusion weighted imaging
EP	end point distance
EPI	echo planar imaging
FA	fractional anisotropy
FDR	false discovery rate
FDT	fiber bundle-driven technique
Fmaj	forceps major
Fmin	forceps minor (frontal projection of the corpus callosum)
GB	gigabyte
GPU	graphics processing unit
HAC	hierarchical agglomerative clustering
HD	Hausdorff distance
HF	high frequency
IFO	inferior fronto-occipital fasciculus
IRQ	inter-quartile range
IS	inferior-superior
LAPACK	linear algebra package
LOF	local outlier factor
LR	left-right
MD	minimum tract distance
MIMD	multiple instruction streams, multiple data streams
MISD	multiple instruction streams, single data stream
MPD	matched point distance
MRI	magnetic resonance imaging
NMR	nuclear magnetic resonance
ODF	orientation distribution function
PGSE	pulsed gradient spin echo
RD	radial diffusion
RGB	RGB (red green blue) color model
ROI	region of interest
SIMD	single instruction stream, multiple data streams
SIMT	single instruction stream, multiple threads
SISD	single instruction stream, single data stream
SLFt	temporal part of the superior longitudinal fasciculus

SMP	symmetric multiprocessor
TBSS	tract-based spatial statistics
TCD	tract centroid distance
TOS	tract orientation similarity
UNC	uncinate fasciculus
VBM	voxel-based morphometry
WM	white matter

Contents

Acknowledgment	I
Zusammenfassung / Abstract	III
Acronyms and abbreviations	VII
Contents	XI

I Introduction and background

1	Introduction	3
1.1	Origin of the brain	3
1.2	A short history of brain research	5
1.3	Motivation	7
1.4	Objectives and outline of the thesis	10
2	Fundamentals	13
2.1	Brownian motion and diffusion	13
2.2	Diffusion in the brain	16
2.3	Probing diffusion with <i>magnetic resonance imaging</i> (MRI)	20
2.3.1	MRI basics	20
2.3.2	<i>Diffusion weighted imaging</i>	22
2.3.3	<i>Diffusion tensor imaging</i>	25
2.4	Tractography	27
2.5	Quantitative analysis of diffusion MRI data	30
2.6	Cluster analysis and similarity measures	32
2.7	Parallel computing	35
2.7.1	General definitions	35
2.7.2	Tasks, processes and threads	36
2.7.3	GPU computing and CUDA	37

II Cluster analysis of fiber tracts

3	Cluster analysis of large tractography datasets	41
---	---	----

Contents

3.1	Introduction	41
3.2	<i>Cluster analysis through smartly extracted representatives (CATSER)</i>	43
3.2.1	Overview	43
3.2.2	Basic CATSER workflow	44
3.2.3	Formation of clusters	45
3.2.4	Outlier handling	47
3.2.5	<i>Local outlier factors</i>	48
3.2.6	Choosing the number of representatives	51
3.2.7	Labeling and reassigning tracts	53
3.2.8	Parameter selection	54
3.2.9	Differences to CURE	56
3.3	Integration of anatomical information into the cluster analysis	57
3.3.1	The white matter atlas	58
3.3.2	Determining the class membership for tracts	58
3.3.3	Determining the weighting factor for two clusters	59
4	Similarity measures for fiber tracts	63
4.1	Overview	63
4.2	<i>Tract centroid distance measure</i>	64
4.3	<i>Tract orientation similarity measure</i>	65
4.4	<i>Minimum tract distance</i>	65
4.5	<i>Hausdorff distance measure</i>	65
4.6	<i>Matched point distance</i>	66
4.7	<i>Combined distance measure</i>	66
5	Technical aspects and implementation details	73
5.1	General implementation details	73
5.1.1	Extracting a random sample of tracts	74
5.1.2	Performing the clustering	75
5.1.3	Assigning remaining objects	75
5.2	Mechanisms for the computation of similarity measures	75
5.2.1	Multithreaded computation of similarity measures	76
5.2.2	Reusing results of previously computed features and similarity measures	78
5.2.3	Processing and processing order	79
5.2.4	Memory management	79
5.3	Algorithms for the computation of the Hausdorff distance	82
6	Clustering experiments	87
6.1	Data acquisition and data processing	87
6.2	Creation of a white matter atlas	88
6.3	Spatial matching of clusters and atlas classes	90
6.4	Experiments	91
6.4.1	Cluster analysis	91
6.4.2	Outlier elimination	92

6.4.3	Performance analysis	93
6.5	Results	94
6.5.1	Cluster analysis	94
6.5.2	Outlier elimination	99
6.5.3	Performance analysis	101

III Quantitative analysis

7	Quantitative analysis of fiber bundles	109
7.1	Introduction	109
7.2	Quantitative analysis of fiber bundles	110
7.2.1	General workflow	110
7.2.2	Projection of quantitative values	114
7.2.3	Spatial normalization and separation of fiber bundles	115
7.2.4	Gridding of quantitative values	115
7.2.5	Voxel-wise statistical analysis of individual fiber bundles	116
7.3	Experiments	117
7.3.1	Dependency of the resolution, number of tracts and distances between tract points	118
7.3.2	Investigation of normalization-induced interpolation	119
7.4	Results	121
7.4.1	Dependency of the resolution, number of tracts and distances between tract points	121
7.4.2	Investigation of normalization-induced interpolation	123
8	Applications	129
8.1	Assessing hemispheric white matter differences in healthy volunteers	129
8.1.1	Introduction	129
8.1.2	Data acquisition, data processing and statistical analysis	130
8.1.3	Results	131
8.2	Assessing hemispheric white matter differences in schizophrenic patients . .	136
8.2.1	Introduction	136
8.2.2	Data acquisition, data processing and statistical analysis	137
8.2.3	Results	137

IV Discussion and conclusion

9	Discussion and conclusion	145
9.1	Cluster analysis	146
9.2	Quantitative analysis of fiber bundles	150
9.3	Applications	154
9.4	Conclusion	156

Contents

Bibliography	XV
List of Figures	XXXI
List of Tables	XXXV
V Appendix	
Publications	XXXIX
Ehrenwörtliche Erklärung	XLV

Part I

Introduction & background

1 Introduction

Brain: An apparatus with which we think what we think.

— Ambrose Bierce (1842–1914)
The Devil's Dictionary

1.1 Origin of the brain

The brain as the center of the *central nervous system* (CNS) is the most complex organ of the human body. With a highly specialized network of billions of interconnected nerve cells (i.e., neurons) that transmit electrochemical signals simultaneously, it regulates autonomic processes and sustains vital body functions. By processing complex sensory inputs, the CNS is able to adapt to changing environments, coordinate complex behavior and handle cognitive tasks.

The evolution of the brain into a complex and highly organized structure as we know it took billions of years and started with the evolution of life itself. One-celled organisms as the simplest form of life exist autonomously and were presumably the first organisms that have inhabited the earth. Even though they can adapt to changing environments, reproduce and extract nutrition to maintain their metabolism, evolution led to a transition from independent single-celled organisms to agglomerations of independent cells that finally transformed into multicellular organisms (Alberts et al., 1994; Ratcliff et al., 2012).

This transition to a more complex organism was characterized by collaboration and by division of labor (Alberts et al., 1994). As cells specialized and cooperated, they formed a new, individual and more complex organism, which was more than just the sum of its parts. With the ongoing specialization and increasing cell differentiation, the complexity of the organism increases, which favored the evolution of elaborate mechanisms for signal propagation. It is hypothesized that the first neurons with fast electrical signal propagation were controlling animal locomotion (Jékely, 2011). One of the simplest neural networks can be found in Cnidaria (Kaiser and Varier, 2011). Even though it possesses only a diffuse two-dimensional nerve net with minimal centralization, it consists of a set of interconnected neurons that are the basic building blocks for more complex nervous systems. However, the neural network

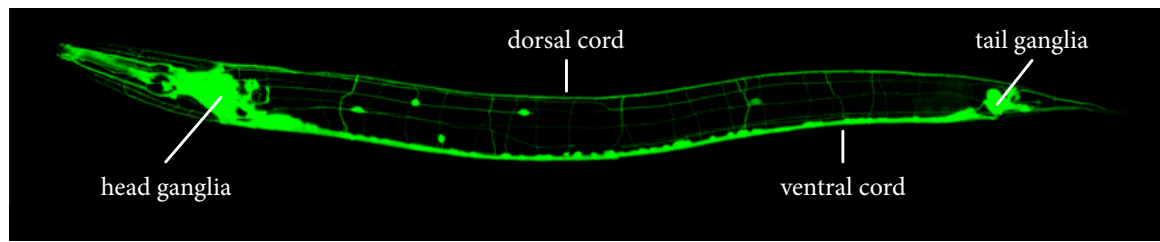


Figure 1.1: Visualization of the nervous system of *Caenorhabditis elegans* using *green fluorescent protein* (modified from Hutter (2008)). Green dots depict neurons while the fine lines show the axonal connections between the neurons.

of Cnidaria is mostly decentralized and has only limited capabilities. For complex behavior such as sensory and motor functions, functional specialization is required, which begins with spatial aggregation of neurons and leads to more complex neural networks, also known as ganglia (Kaiser and Varier, 2011).

One of the most studied organisms is the transparent worm *Caenorhabditis elegans* (usually referred to as *C. elegans*), whose genome was fully sequenced by Brenner (1974). *C. elegans* has a precisely timed development cycle and evolves into a hermaphrodite adult worm that possesses 959 body cells (Alberts et al., 2008). 302 of these cells are nerve cells, located in two ganglia at the front and the back that are interconnected by a neural cord and a subset of smaller ganglia. Since its anatomy has been described in extraordinary detail and the structure of the nervous system has been thoroughly mapped, including practically all neural connections (White et al., 1986), it is the perfect model organism to study neural processing, neural function and neural development. The nervous system of the worm including its neurons and their connections (i.e., axons) is shown in Figure 1.1. Possessing only 302 neurons and a simple neural network, *C. elegans* is capable to perform the most fundamental functions (movement, nutrition, reproduction, et cetera).

In more advanced organisms, the ganglion has further evolved into a complex network of highly interconnected neurons with a centralized, hierarchical structure – the brain. Even though architecture changes and cerebral evolution among species diverges¹, the basic structure and fundamental function of the brain is conserved while the brain itself specializes and adapts to the environment (Nicholson, 2001; Sarnat and Netsky, 2002). The evolutionary process of the brain is not only characterized by the immanent centralization and a high specialization of areas that are dedicated to important functions of the species (e.g., visual and auditory perception, motor control or speech), but is also accompanied by the synthesis of a complex network with a high degree of connectivity between neurons of the same area that are linked to distant areas and form signal pathways in order to enable long-range signal propagation. The formation of a highly specialized network with billions of interconnected neurons and distinct processing centers seems to be the key ingredient that gives rise to higher brain functions.

¹For example, neurons in reptiles and birds form clusters while they form layers in mammals.

1.2 A short history of brain research

With the evolution of the human species and the rise of civilizations, culture and science began to flourish and knowledge increased. It was in old Egypt, where the *Edwin Smith surgical papyrus*, the first written document in recorded history about medicine, referenced the brain and the nervous system (see Figure 1.2). Even though the text originated around 1700 B.C., it is only a replica of a much older surgical treatise that dates back even further to about 3000 B.C. (Gross, 1999).

In ancient Greece, around 400–500 B.C., the function and the purpose of the brain were not well understood and various theories existed where human intellect arises. No theory was universally accepted but two opinions were prevalent and attributed the origin of intellect to either brain or heart. While Alcmaeon of Croton (around 450 B.C.), who seems to be the first practitioner of dissection, proclaimed that the brain was the site of sensation and cognition, Empedocles of Agrigento (ca. 490–430 B.C.) thought that the heart was the central organ of intellect and that the composition of the blood influences intelligence (Gross, 1999). Both views had influential followers. The Hippocratic doctors of ancient Greece supported the first hypothesis and appointed intellect and mental processes to the brain, while Aristotle considered the brain only to be a secondary organ that tames the animal beast by cooling its blood (Findlen, 1998).

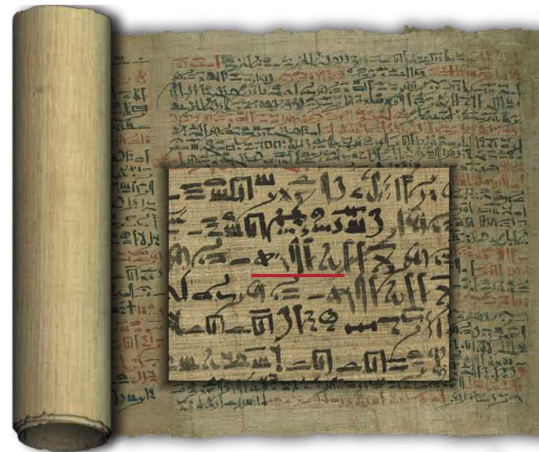



Figure 1.2: First occurrence of the word  (i.e., brain, underlined in red) in the digitized version of the *Edwin Smith surgical papyrus* (modified from the U. S. National Library of Medicine (2012)).

In Alexandria of the second century, Herophilus of Chalcedon (335–280 B.C.) and Erasistratus of Ceos (ca. 304–250 B.C.) began the first systematic study of human anatomy, carried out not only by dissection of human cadavers but also by vivisection of living criminals (Gross, 1999). While Herophilus and Erasistratus followed the idea that the brain is the center of sensation and cognition, they were particularly interested in studying its function and were the first to provide detailed and accurate descriptions of the human brain, the ventricular system and the connectivity of certain nerves (Gross, 1999).

During the time of the Roman Empire, the dissection of the human body was forbidden by law, which led the Roman physician Galen of Pergamon (129–199) to focus his studies on animals. In his studies, Galen dissected and vivisected not only animals including pigs, oxen and elephants, but also primates as he assumed a good resemblance to the human anatomy (Guerrini, 2003). These animal studies but also his extensive clinical experience in the gladiator school of Pergamon, enabled him to give an extraordinarily detailed and accurate description of brain and nerve anatomy, which endured throughout history and influenced physicians until the Renaissance (Gross, 1999).

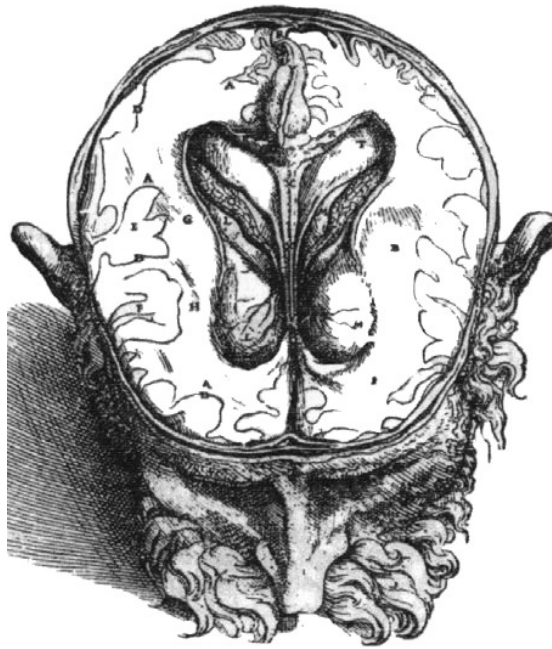


Figure 1.3: Drawing of the brain and the ventricular system in the transverse plane from Vesalius's *De homini corporis fabrica* (1543). Two distinct compartments in the brain are clearly delineated (from Tamraz and Comair (2005)).



Figure 1.4: Drawing of a coronal section through the brain that shows a *body with medullary streaks* (i.e., corpus striatum). The difference between grey and white matter is particularly well illustrated in the cerebellum (by Willis (1664), from Catani and Thiebaut de Schotten (2012)).

With the death of Galen and the decline of the Roman Empire in the 5th century, the progress of medical sciences stopped until the eagerness for knowledge was rejuvenated with the beginning of the *Medical Renaissance*, 1000 years later. The re-establishment of human dissection in 1316 by Mondino dei Luzzi (ca. 1270–1326) and his treatise on anatomical dissection *Anathomia corporis humani* (first printed around 1478), renewed the medical sciences and influenced the work of succeeding anatomists. In the 16th century medicine and anatomy began to flourish. Andreas Vesalius (1514–1564) published his monumental seven volume masterpiece *De homini corporis fabrica* (1543), one of the most enduring and influential text books in the history of science (Filley, 2001). Besides his contributions to medicine and human anatomy, he was also the first to recognize two distinct compartments in the brain. Although Vesalius delineated two neuroarchitectural compartments in his drawings (see Figure 1.3), he never addressed the separated structures in his writings (Catani and Thiebaut de Schotten, 2012). Soon after the publication of the *fabrica*, Archiangelo Piccolomini (1526–1586) made the first clear distinction between the two neuroanatomical structures that he termed *cerebrum* (cerebral cortex or grey matter) and *medulla* (white matter) (Filley, 2001). His observations marked the beginning of the anatomical study of the cerebral cortex and the white matter (Catani and Thiebaut de Schotten, 2012). Thomas Willis (1621–1675) for example, made an early attempt to describe the function of white matter and speculated

that it elaborates sensory signals into perceptions and imaginations that were later stored as memories in the cortex (Filley, 2001; Gross, 1999). One of Willis's drawings, a coronal section through the brain that illustrates the white and the grey matter is shown in Figure 1.4.

The invention of microscopy accelerated the scientific revolution. Marcello Malpighi (1628–1694), who was the first to use the new technique for the examination of the white matter architecture, discovered that the roots of the nerve fibers either terminate or originate in the cortex and that the *white medullary substance* (as he called it, i.e., white matter) of the brain can be attributed to a multitude of these connections (Gross, 1999). Based on Malpighi's observations, Emanuel Swedenborg (1688–1772) developed a comprehensive theory for the structural and functional organization of the cortex and the white matter (Gross, 1999). For example, he correctly assigned sensory and motor functions to the cortex and concluded that these functional areas are connected to the body by nerves that travel through the brainstem along the spinal cord. He also stated that the functionally independent units of the brain are connected by fibers and that the corpus callosum interconnects both hemispheres (Gordh et al., 2007). However, Swedenborg's writings were barely noticed and it took scientists over a century to draw the same conclusions that finally confirmed Swedenborg's theories.

With the discovery of the *black reaction* – a silver nitrate staining technique for nervous tissue in 1873, Camillo Golgi (1843–1926) introduced a novel way to study the architectural organization of the brain. Golgi's method was later used by Santiago Ramón y Cajal (1852–1934), who finally revealed the complex microstructure of brain tissue with its interconnected neurons. An example of Ramón y Cajal works is shown in Figure 1.5.

After 5000 years of brain research and fundamental discoveries, the basic structure of the brain and the nervous system was finally revealed. While the *neuron doctrine* constituted the beginning of modern neuroscience, the functional and structural analysis of nervous tissue just started.

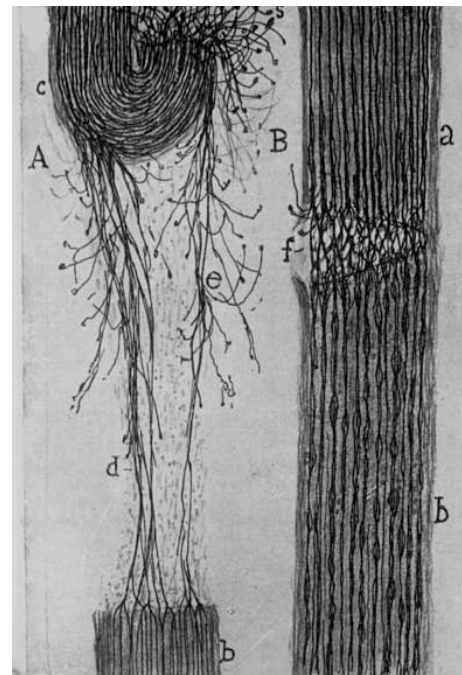
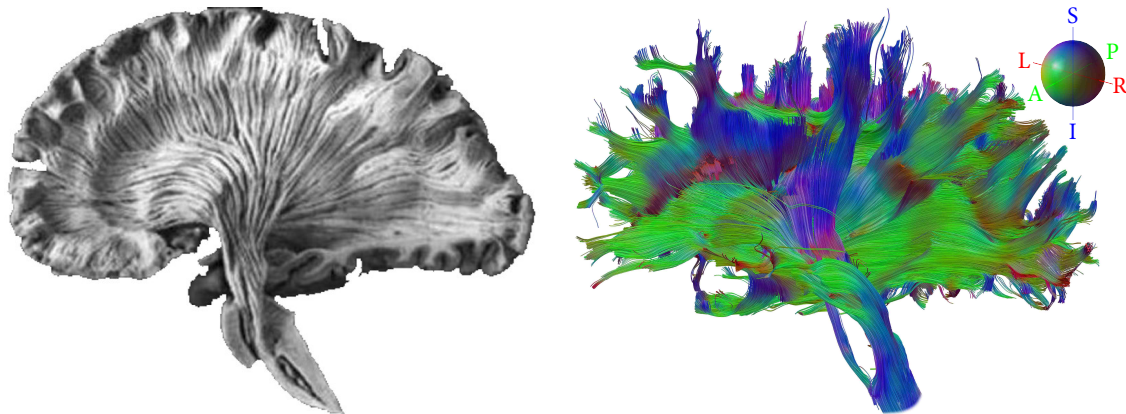


Figure 1.5: Drawing by Ramón y Cajal (1913) who depicted the regeneration of a nerve fiber after hemisection.

1.3 Motivation

Every living organism uses diffusion as a mechanism to transport and exchange substances. The human brain as the neural processing center consists of a highly complex cellular microstructure in which diffusion plays a crucial role for brain function (Nicholson, 2001). Diffusion is a fundamental process and highly depends on the microstructure and the cellular organization of the tissue. The brain primarily consists of interconnected nerve cells and can be divided into two distinct compartments – the grey matter with the bodies of the nerve



(a) Photograph of a *post mortem* brain specimen. (b) 3D reconstruction of *in vivo* fiber tracts.

Figure 1.6: A set of white matter fiber tracts. (a) shows selected tracts of one hemisphere in a *post mortem* brain specimen after dissection and special preservation by Williams et al. (1999). (b) shows tracts in a healthy subjects obtained with *diffusion tensor imaging* (DTI) and fiber tractography. Fiber tracts are color-coded according to their spatial orientation with red = left-right, green = anterior-posterior and blue = inferior-superior.

cells and the white matter where their axonal connections are concentrated. The axons in the white matter are coherently aligned and are often organized in fiber bundles that interconnect different functional areas of the brain (see Figure 1.5 for an example of a fiber bundle). With their dense packing and their axonal membranes, they form highly orientated barriers. While molecules experience less resistance parallel to the axons, they are hindered perpendicular to the nerve fibers and diffuse in the direction parallel to the axons. The diffusion in the white matter becomes directionally dependent and diffusion anisotropy arises.

With the invention of *magnetic resonance imaging* (MRI) by Paul C. Lauterbur (1929–2007) and Sir Peter Mansfield (1933) in the 1970s and the introduction of *diffusion weighted imaging* (DWI) by Le Bihan and Breton (1985) as well as *diffusion tensor imaging* (DTI) by Basser et al. (1994), the quantification of diffusion became possible. By using DTI, tomographic, cross-sectional images can be reconstructed that characterize the diffusion in each imaging voxel with a second order diffusion tensor. This facilitated the investigation of the unique structural organization of nervous tissue and diffusion anisotropy in the brain *in vivo*.

As the resulting voxel-wise diffusivity profiles of the diffusion tensor are potential indicators for the underlying microstructural axonal pathways in the brain (Leergaard et al., 2010), tractography techniques (Basser et al., 2000; Conturo et al., 1999; Mori et al., 1999; Parker and Alexander, 2005; Lazar and Alexander, 2005; Behrens et al., 2007) can be used to approximate the underlying white matter structures with three-dimensional fiber trajectories (Mori and van Zijl, 2002; Dauguet et al., 2007). An example for some white matter fibers in an *ex vivo* brain specimen and *in vivo* fiber tracts reconstructed from a DTI dataset of a healthy volunteer are shown in Figure 1.6. For whole brain tractography, the reconstructed datasets contain a wealth of information and consist of several thousand up to more than one million streamlines. Though such datasets approximate the underlying brain structure in high detail, the fiber

tracts (i.e., streamlines) have no apparent structural organization and are loosely distributed throughout the brain. It remains thus unclear to which underlying white matter structure particular fiber tracts belong and if tracts are part of either the same or of distinct structures.

Fiber tracts are often color-coded according to their spatial orientation (see Figure 1.6b). This coloring is mainly a visual aid that does not help to decipher the structural organization of the tractography datasets. However, various potentially useful applications would greatly benefit from disentangling the complex fiber network. Fiber tracts, grouped into meaningful bundles that represent the underlying white matter structures correctly, are useful to assess structural connectivity between distinct brain regions (Lanyon, 2012) or to determine structural integrity of distinct white matter pathways.

Correct assignment of fiber bundles would also be helpful for the assessment of tumors and the delineation of tumorous tissue, as this will aid to determine if white matter bundles have been infiltrated by the tumor or whether the bundles have merely been dislocated (Stadlbauer et al., 2007; Nucifora et al., 2007). The incorporation of such fiber bundle specific information (e.g., course, spatial location, integrity) into treatment planning, neuronavigation as well as radiation therapy will aid the neurosurgeon and ultimately help the patient.

From the diffusion tensor, various quantitative measures can be extracted that characterize distinct properties of diffusion. As alterations in the diffusivity are often indicators for structural changes that are related to either natural or pathological causes (Pandit, 2009), DTI is well suited to study disorders that are associated with changes in brain diffusivity. Techniques for quantitative analysis can be applied to assess altered diffusion and to detect statistically significant differences between two or more groups of subjects (e.g., patients vs. healthy controls) or between subgroups of patients. Compared to established and predominantly applied techniques such as *voxel-based morphometry* – VBM (Ashburner and Friston, 2000) or *tract-based spatial statistics* – TBSS (Smith et al., 2006), new *fiber bundle-driven techniques* (FDTs) enable the analysis of individual white matter bundles (Fillard et al., 2003; Gong et al., 2005; O'Donnell et al., 2007). While FDTs are promising for the quantification of group differences in individual bundles, they suffer from interpolation effects between distinct white matter structures that result from the normalization and coregistration of different datasets (Chao et al., 2009).

Resolving the structural organization of tractography datasets is of extraordinary importance for a number of potentially useful applications but lacks applicability as the grouping of fiber tracts into meaningful bundles is difficult. Even though the bundling of fiber tracts can be performed manually, this type of processing is prone to errors, remains highly time-consuming and requires an operator with fundamental neuroanatomical knowledge.

Machine learning methods are auspicious techniques for the automatic extraction of fiber bundles. Classification, for example, is a supervised machine learning method that uses predefined prototype classes (e.g., a white matter parcellation, atlas) to predict the membership of fiber tracts to a class. With increasing availability of atlases and parcellations (Wakana et al., 2004; Verhoeven et al., 2010; Zhang et al., 2010) as well as guidelines to accomplish a reproducible segmentation of the white matter (Wakana et al., 2007), atlas-based classification has become a convenient tool to define fiber bundles that correspond to specific regions of the atlas.

If an atlas can not be used or is not available, fully automated unsupervised learning techniques can be applied instead of supervised methods. Fiber clustering belongs to these unsupervised methods that analyze the similarities between fiber tracts in order to assemble similar tracts into distinguishable fiber bundles. While classification is only able to define fiber bundles that correspond to classes in the anatomical atlas, cluster analysis groups tracts into fiber bundles based on the similarity of distinct features of the tracts. Due to the high variability of different datasets, a consistent, reproducible and anatomically correct extraction of fiber bundles across multiple subjects solely based on tract similarity and without anatomical guidance is difficult to achieve.

One fundamental drawback of clustering is the high computational complexity, which is immanent to the majority of conventional clustering algorithms (Xu and Wunsch, 2008). Since fiber tracts are sets of points that constitute complex trajectories in three-dimensional space, appropriate measures are indispensable to determine the similarity between the fibers. However, both the costly clustering and the complex similarity measures increase the total computational load and typically restrict cluster analysis to small datasets that consist of only a few thousand fiber tracts. Despite the multitude of available methods that has been proposed for both classification and clustering of fiber tracts (Ding et al., 2003; Brun et al., 2004; Moberts et al., 2005; Huang et al., 2005; O'Donnell et al., 2006; Zhang et al., 2008; Maddah et al., 2008; Li et al., 2010; Garyfallidis et al., 2012; Wu et al., 2012; Wang et al., 2013), fast and consistent clustering for multiple subjects is still challenging. While this hinders the automatic extraction of bundles, it prevents not only the analysis of the anatomy and the structural integrity of neural fiber bundles but also severely limits quantitative analysis in group-based imaging studies.

1.4 Objectives and outline of the thesis

As the applicability of fiber tract clustering is seriously limited with recent methods, the main objective of this thesis was to develop a cluster analysis framework that enables fast clustering of large tractography datasets in reasonable time. In order to facilitate the study of alterations in fiber bundles, the second intent of the present work was to introduce a new technique for fiber bundle-based quantitative analysis that is applicable in group-based imaging studies. By using various experiments and initial studies, the thesis demonstrates the practicability of the presented techniques.

To emphasize the outline and the contributions of this work, the thesis is organized in five distinct parts. Beginning with part I, an introduction and motivation for the thesis is presented (chapter 1). Fundamentals and basic concepts that are relevant for this thesis are presented in chapter 2. This chapter introduces not only the physics of diffusion, but also the origin of diffusion in tissue and expounds how anisotropic diffusion arises in the white matter of the brain. The principles of *magnetic resonance imaging*, *diffusion weighted imaging* and fiber tractography are described. An overview of existing methods for quantitative analysis in group-based imaging studies is presented before the chapter closes with a brief introduction on cluster analysis and parallel computing.

Part II is dedicated to the cluster analysis of fiber tracts. CATSER (*cluster analysis through smartly extracted representatives*), a new cluster analysis algorithm is presented as well as a method for the integration of anatomical information into the cluster analysis (chapter 3). In chapter 4, similarity measures that are relevant for the thesis are formally defined and reviewed before important mechanisms for the fast processing of large datasets are described and technical considerations and details for the implementation of the cluster analysis framework in a parallel computing environment are unraveled (chapter 5). To show the applicability of the clustering in a multi-subject study and to demonstrate the computational efficiency of the clustering framework, various experiments have been conducted that are presented in chapter 6.

The third part deals with the quantitative analysis of extracted fiber bundles. While a new method for quantitative analysis of fiber bundles is presented in chapter 7, its application to group-based analysis is described in chapter 8. To investigate the applicability of the presented methods, two studies were conducted to explore hemispheric diffusion differences of selected fiber bundles in healthy volunteers and schizophrenic patients.

In part IV, the presented techniques and results are discussed and a conclusion and an outlook is given (chapter 9). The thesis closes with part V which contains an appendix with additional information.

2 Fundamentals

*The more you know, the more you realize
you know nothing.*

— Socrates (469 B.C.–399 B.C.)

In this chapter, the fundamental principles of Brownian motion and diffusion are reviewed as well as the origin of diffusion contrast in the brain. After a short introduction to *magnetic resonance imaging* (MRI), *diffusion weighted imaging* (DWI) and *diffusion tensor imaging* (DTI), the basics of fiber tractography and quantitative analysis of DWI data are presented. To conclude this chapter, essential basics of cluster analysis and similarity measures as well as parallel computing are reviewed.

The foundations of MRI are only touched briefly as they are not essential to understand the thesis. For a more detailed, technical introduction to this subject the reader is referred to the books by Haacke et al. (1999); Levitt (2001) and Hendee and Ritenour (2002). A comprehensive description to the underlying principles of quantum mechanics is given by Cohen-Tannoudji et al. (2006). Even though the principles of *diffusion weighted imaging* and *diffusion tensor imaging* are partly covered, the interested reader may like to take a closer look into the book by Jones (2011) that exhaustively covers everything related to diffusion MRI. For further information on cluster analysis and similarity measures the comprehensive books of Xu and Wunsch (2008) and Everitt et al. (2011) are advised.

2.1 Brownian motion and diffusion

Diffusion is the translational movement of particles that occurs as a result of random molecular motion in the absence of bulk motion. If concentration differences exist in a substance, a flow of particles is observed and particles from areas with high concentration will diffuse to areas with lower concentration, which will result in a final uniform distribution of the substance. On a macroscopic scale, the flux of particles $\mathbf{J}(\mathbf{r}, t)$ can be described with Fick's first law (Fick, 1855)

$$\mathbf{J}(\mathbf{r}, t) = -D\nabla\phi(\mathbf{r}), \quad (2.1)$$

where D is the diffusion constant and $\nabla\phi(\mathbf{r})$ the gradient of concentration $\phi(\mathbf{r})$ at position \mathbf{r} (Crank, 1975). The diffusion constant D is also called diffusivity and determines the mobility of the diffusing substance. The flux \mathbf{J} is a mass flow rate that quantifies the number of particles that pass through an area in a specific amount of time. Fick's first law only describes steady-state diffusion under the assumption that the concentration gradient is constant over time. In practice, however, the concentration and the concentration gradient change with time and non-steady-state diffusion occur. In order to describe the change of concentration over time, additional aspects have to be considered. The law of conservation of mass requires that in a closed system, the total number of particles has to be maintained. The number of particles that diffuse into an area must be equal to the sum of particles that flow out of the surrounding area. This is mathematically expressed by the continuity equation (Ursell, 2011)

$$\frac{\partial\phi(\mathbf{r}, t)}{\partial t} + \nabla\mathbf{J}(\mathbf{r}, t) = 0. \quad (2.2)$$

Inserting Eq. 2.1 into Eq. 2.2 yields the linear partial differential diffusion equation (Fick's second law) that describes the change of the concentration with time:

$$\frac{\partial\phi(\mathbf{r}, t)}{\partial t} = \nabla(D\nabla\phi(\mathbf{r}, t)). \quad (2.3)$$

If D is constant in space, it follows according to Ursell (2011)

$$\frac{\partial\phi(\mathbf{r}, t)}{\partial t} = D\nabla^2\phi(\mathbf{r}, t). \quad (2.4)$$

Even though Fick's laws were derived for the transportation of particles through a diffusion gradient, diffusion also occurs in the absence of such a concentration gradient, which is then called self-diffusion. The underlying physical phenomenon for diffusion was first discovered by Brown (1828) and is thus called Brownian motion. For studies on pollen, Brown suspended pollen grains in water and observed ceaseless, random motion of the grains using a microscope (Ghosh, 2009). Brown was never able to explain his observation and more than half a century passed before Einstein (1905) derived the mathematical formalism that relates the random movement of particles to thermal molecular motion and thus explains the Brownian motion (Perrin, 1909). In essence, particles, irrespective of their current state, exhibit random motion if they possess thermal energy and the temperature is above absolute zero (0 K). The movement of the pollen grains can thus be explained by the random motion of water molecules that constantly collide with the grains. The collisions occur from different directions but are not able to neutralize each other, resulting in the random movement of the pollen grain. Consequently, self-diffusion is the translational displacement of particles due to the Brownian motion.

Fick's laws are macroscopic descriptions of diffusion that is governed by the Brownian motion on a microscopic scale. The mathematical formalization of the Brownian motion enabled its interpretation as a continuous-time stochastic process, in which the probability densities obey the differential equations (Callaghan, 2011). Let us assume that we have a specific amount of particles M that are added as a Dirac function at the origin $\mathbf{r} = \mathbf{0}$ and the

initial time point $t = 0$ (Callaghan, 2011). If the substance is isotropic and the diffusivity is directionally independent, the solution to the diffusion equation (Eq. 2.4) for three dimensions is given according to Einstein (1905) and Ursell (2011) as

$$\phi(\mathbf{r}, t) = \frac{M}{(4\pi Dt)^{3/2}} \exp\left(-\frac{\mathbf{r}^2}{4Dt}\right). \quad (2.5)$$

From Eq. 2.5 it follows that the distribution of the particles is Gaussian and the width of the distribution grows with time. Let us assume that $P(\mathbf{r}|\mathbf{r}_0, t)$ is the conditional probability that a particle that was at the initial spatial location \mathbf{r}_0 at time $t = 0$ will be at location \mathbf{r} after time t . To obtain the time-dependent probability for the displacement, the diffusion equation (Eqs. 2.3 and 2.4) is rewritten as (Callaghan, 2011)

$$\frac{\partial P(\mathbf{r}|\mathbf{r}_0, t)}{\partial t} = \nabla \cdot (D \nabla P(\mathbf{r}|\mathbf{r}_0, t)), \quad (2.6)$$

$$\frac{\partial P(\mathbf{r}|\mathbf{r}_0, t)}{\partial t} = D \nabla^2 P(\mathbf{r}|\mathbf{r}_0, t). \quad (2.7)$$

The solution for Eq. 2.7 follows from Eq. 2.5 and is the probability distribution (Callaghan, 2011)

$$P(\mathbf{r}|\mathbf{r}_0, t) = \frac{1}{(4\pi Dt)^{3/2}} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}_0|^2}{4Dt}\right). \quad (2.8)$$

The mean of this distribution is $\mu = \mathbf{r}_0$ and the variance $\sigma^2 = 2Dt$, which is the average mean-squared displacement for all particles.

Eq. 2.7 is only valid for diffusion in isotropic media (Callaghan, 2011). In anisotropic media, the diffusion is orientation-dependent and the diffusion properties vary in different directions. In such media, the diffusing particles have a preferential direction (Crank, 1975) and a single scalar value D is inadequate to describe the diffusion process accurately. Instead of using the scalar diffusion coefficient D , more general models have to be used in order to characterize the orientation dependency. In anisotropic media, the diffusion can be approximated with a 3×3 second order, symmetric and positive definite diffusion tensor $\underline{\mathbf{D}}$

$$\underline{\mathbf{D}} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}. \quad (2.9)$$

The differential equation (Eq. 2.7) changes to (Callaghan, 2011)

$$\frac{\partial P(\mathbf{r}|\mathbf{r}_0, t)}{\partial t} = \nabla \cdot (\underline{\mathbf{D}} \nabla P(\mathbf{r}_0|\mathbf{r}, t)), \quad (2.10)$$

which yields the direction-dependent, conditional probability of translational displacement for the diffusing particles. By using the diffusion tensor, the probability distribution (Eq. 2.8) evolves according to Bassler et al. (1994)

$$\frac{\partial P(\mathbf{r}|\mathbf{r}_o, t)}{\partial t} = \frac{1}{\sqrt{\det(\underline{\mathbf{D}})(4\pi t)^3}} \exp\left(-\frac{(|\mathbf{r} - \mathbf{r}_o|^2)^T \underline{\mathbf{D}}^{-1} (|\mathbf{r} - \mathbf{r}_o|^2)}{4t}\right). \quad (2.11)$$

Eq. 2.11 describes the evolution of the distribution probability for diffusing particles in anisotropic media over time. At each time point t , this distribution probability can be represented with a time-dependent three-dimensional diffusion ellipsoid, whose size, shape and orientation correspond to important properties of anisotropic Gaussian diffusion (Basser and Özarslan, 2011).

In order to visualize the mean-squared displacement of the diffusing particles, the symmetric diffusion tensor $\underline{\mathbf{D}}$ can be diagonalized

$$\underline{\mathbf{D}} = \mathbf{V}^{-1} \boldsymbol{\lambda} \mathbf{V}, \quad (2.12)$$

with eigenvectors \mathbf{V} and corresponding eigenvalues $\boldsymbol{\lambda}$

$$\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3] = \begin{bmatrix} v_{x1} & v_{x2} & v_{x3} \\ v_{y1} & v_{y2} & v_{y3} \\ v_{z1} & v_{z2} & v_{z3} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}. \quad (2.13)$$

This yields the ellipsoidal representation of the diffusion tensor, where the three eigenvectors define the principal directions of the diffusion and the corresponding eigenvalues the principal diffusivities (Basser et al., 1994) (see Figure 2.1a). Generally, eigenvalues and the corresponding eigenvectors are sorted according to the strength of the diffusion with $\lambda_1 > \lambda_2 > \lambda_3$. The dominant direction of diffusion is then indicated by the first eigenvector and the corresponding eigenvalue. In the special case when diffusion is isotropic, the ellipsoid is spherical (see Figure 2.1b) and the diffusion tensor reduces to

$$\underline{\mathbf{D}} = D \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2.14)$$

An example for the diffusion tensor ellipsoid in an anisotropic medium is given in Figure 2.1c.

2.2 Diffusion in the brain

In nature, every living organism uses diffusion as a basic mechanism to transport and exchange substances. The microstructure of multicellular organisms is composed of different types of cells with each cell consisting of a membrane that encases the cell and shields its interior components from the outside (Whitford et al., 2011). The membrane acts as a barrier between the cell and its surrounding, and mass exchange between intra- and extracellular space occurs

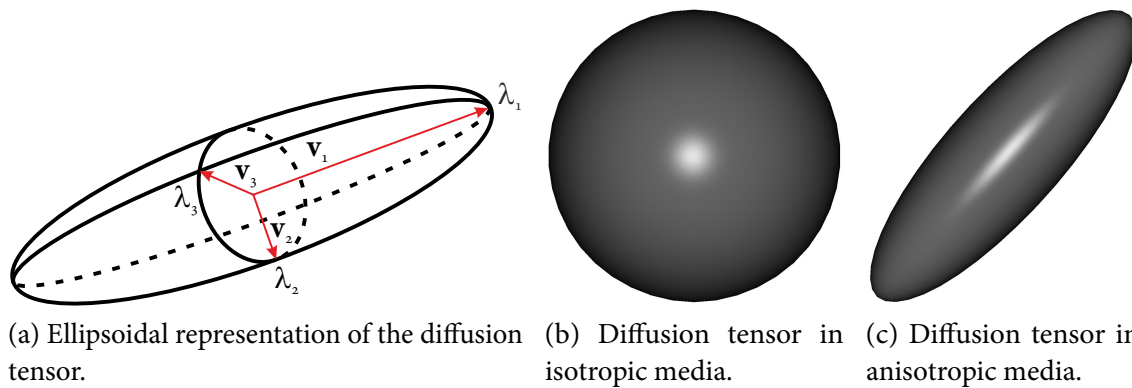


Figure 2.1: Ellipsoidal representation of a diffusion tensor, defined by its eigenvectors \mathbf{V} and eigenvalues λ (a). The diffusion tensor is spherical in isotropic media (b), while it is an ellipsoid that unveils the preferential direction of diffusion in anisotropic media (c).

through active or passive transport mechanisms (e.g., diffusion, osmosis, ion pumps). The underlying microstructure of an organism is highly complex and the cellular organization influences diffusion of molecules and particles in the organism. The probability distribution for the particles is no longer Gaussian due to the different compartments that act as barriers and hinder diffusion (Le Bihan et al., 2001). The overall mobility of diffusing particles in tissue is obviously influenced by the distance between obstacles and particles and the observed time frame of the diffusion process, but also by the configuration of the cellular structures, the distance between obstacles and the number and shape of the obstacles (Callaghan, 1991).

The human brain as the center of the *central nervous system* (CNS) consists of a highly complex cellular microstructure in which diffusion plays a crucial role for brain function (Nicholson, 2001). The brain itself consists of two principal types of cells – nerve cells and glia cells that are surrounded by fluid. Even though cells are densely packed in the CNS, 20 % of the volume fraction in the brain belongs to the extracellular space, filled with *interstitial fluid* that is in contact with *cerebrospinal fluid* at the ventricular surfaces (Syková and Nicholson, 2008). In extracellular space, diffusion is the dominant transport mechanism that is not only responsible for the autoregulation of brain functions, but also for the nutrition of cells that need oxygen, glucose, proteins and other substances to sustain their metabolism (Nicholson, 2001).

Glia cells are scattered throughout the brain and play an integral role in the CNS. While they add structural support and stabilize nerve cells, they also maintain the metabolic function of the brain and regulate synaptic plasticity which is believed to be the basis for learning and memory (Todd et al., 2006; Bélair et al., 2010).

The neural processing of the brain is predominantly attributed to the interconnected network of electrically excitable neurons. In its most simplified version, a neuron consists of the cell body (i.e., soma), the nerve fiber (i.e., axon) and dendrites (see Figure 2.2). The axon is a long projection of the nerve cell and can branch out into a treelike structure that connects to multiple other nerve cells in order to simultaneously transmit the electrochemical signals of the neuron to these other cells. While a neuron can only have one axon, it can possess

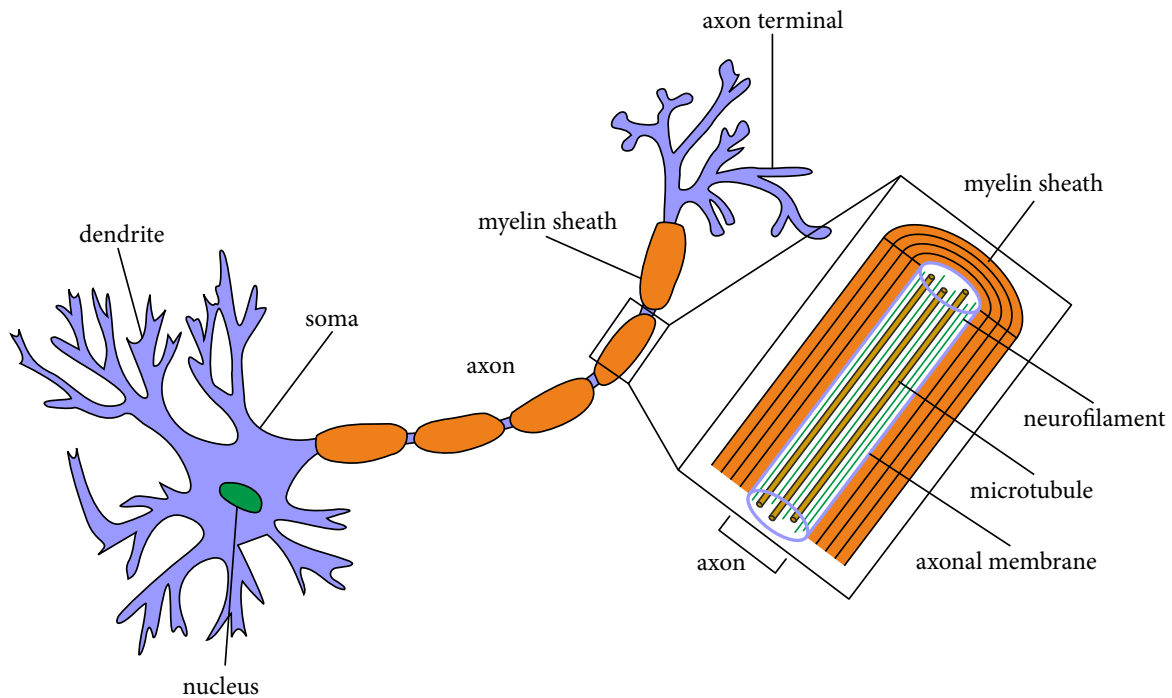


Figure 2.2: Simplified schematic of a neuron, some dendrites and its axon (adapted from Jarosz (2009) and Beaulieu (2002)).

multiple dendrites that are connected to the axons of other nerve cells in order to receive their signals. Thus, the key to neural processing are these interconnected neurons that receive and transmit electrochemical signals from multiple neurons (see Figure 2.3a for an exemplary image that shows pyramidal neurons that form a network).

The neuron bodies are not randomly distributed in the brain, but are spatially located at the surface (i.e., cerebral cortex) and in well delineated areas inside the brain (e.g., brain stem) or the spinal cord. Regions that contain predominantly the bodies of neurons are called grey matter¹. Depending on the region, grey matter consists of multiple layers that have different cellular compositions and distinct distribution patterns of the neuron bodies.

Even though diffusion in the brain is also at work for extrasynaptic neurotransmission as a means of intercellular communication, generally known as volume transmission (Zoli et al., 1998; Fuxe et al., 2012), the primary mechanism for intercellular communication is the axonal wiring and the transmission of electrochemical signals between the neurons². The inner regions of the brain are termed white matter and contain the axons of the neurons, which guide the electrochemical signals to spatially distinct regions of the brain. An essential

¹The term grey matter is attributed to the color of the cell bodies after formalin fixation.

²It was believed that glia cells are not involved in neurotransmission, but Gourine et al. (2010) presented evidence that they at least modulate neurotransmission and actively contribute to information processing of the CNS.

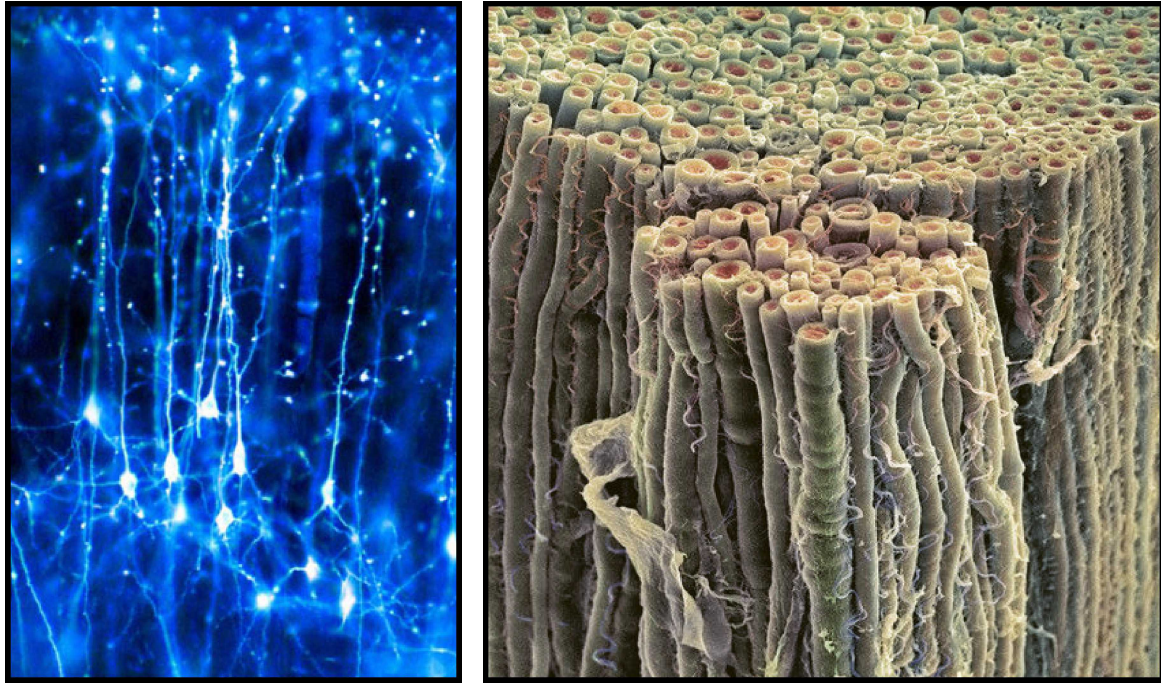
distinctive feature of axons is the axonal myelination³. Myelinated axons are insulated with sheaths of myelin (see Figure 2.2), which increase the nerve conduction velocity and thus the speed of the electrochemical signals that propagate along the nerve fibers. Inside the axon, it contains the cytoskeleton as well as microtubules and neurofilaments.

Since neuron bodies do not exist in the white matter but mostly axons, the structural organization between grey and white matter differs significantly. Even though the number of neurons and their connections greatly varies among species and the size of the brain, Drachman (2005) estimated that the neocortex, which is the outer layer of the human brain, consists of 20 billion neurons (20×10^9), each connected to an average of 7000 neurons. This accumulates to 0.15 quadrillion connections (0.15×10^{15}) or approximately 1 trillion connections per cm^3 (1×10^{12}) (Drachman, 2005).

The axons in the white matter interconnect different brain regions. Due to the extensive number of connections, it is evident the axons are densely packed. However, axons are not randomly distributed, but coherently aligned throughout the brain and often organized in fiber bundles (see Figure 2.3b) that form highly oriented barriers for diffusing particles or molecules. As a result, diffusion is hindered in the direction perpendicular to the nerve fibers and particles experience less resistance parallel to the axons. As they diffuse in the direction that is parallel to the axons, the diffusion in the white matter becomes directionally dependent and anisotropic. Even though it is not yet fully understood whether anisotropy in white matter originates from intra- or inter-axonal diffusion, it is plausible that both compartments contribute to the anisotropy (Le Bihan et al., 2001; Takahashi et al., 2002; Beaulieu, 2002; Assaf and Basser, 2005; Assaf et al., 2008). However, the exact contribution of intra- and inter-axonal diffusion will depend on the time scale diffusion is observed, the axon diameter as well as extracellular spacing between axons. As pointed out by Beaulieu (2002), many experiments support the hypothesis that intact axonal membranes have a strong influence on diffusion anisotropy especially at short diffusion times (Le Bihan et al., 2001; Takahashi et al., 2002). Even though myelination is not essential for anisotropic diffusion, experiments also suggest that the myelin sheaths modulate the degree of anisotropy (Beaulieu, 2002). In order to assess and characterize intra-axonal diffusion, measurements at longer diffusion times have to be performed (Assaf and Basser, 2005; Assaf et al., 2008).

Alterations in the diffusivity of particles in white matter are often an indicator for structural changes that are associated with changes in the volume ratio between intracellular and extracellular space (Le Bihan et al., 2001). These changes might be due to the reorganization of white matter during brain development (Löbel, 2007; Giorgio et al., 2008). They can, however, also be related to pathological conditions that might lead to demyelination, axonal degeneration or ischemia (Beaulieu, 2002).

³The term white matter has originated from the formaldehyde-based preservation of the brain, in which the myelin changes its color to white.



(a) Color-enhanced *light microscopy* image of pyramidal neurons forming a network in the brain (Clarke, 2012).

(b) Disconnected bundle of nerve fibers, acquired with a *scanning electron microscope* (Gschmeissner, 2011). The axons of the individual nerve fibers are shown in orange and insulating myelin sheaths in green.

Figure 2.3: Colored images of pyramidal neurons (a) and a disconnected bundle of axons (b).

2.3 Probing diffusion with *magnetic resonance imaging* (MRI)

2.3.1 MRI basics

Interactions of elementary nuclear particles (see Figure 2.4) with externally applied magnetic fields enable MRI to probe the atomic structure of a sample and to obtain tomographic images *in vivo*. The fundamental principle of *nuclear magnetic resonance* (NMR) and MRI (Bloch, 1946; Purcell et al., 1946; Lauterbur, 1973) is based on quantum mechanics and the existence of a magnetic moment (nuclear spin I) that is immanent to all atomic and subatomic particles with a spin quantum number other than 0.

When atomic nuclei that possess a magnetic dipole moment μ are exposed to an external magnetic field B_o , μ will precess around B_o with the *Larmor frequency* ω_o :

$$\omega_o = \gamma B_o, \quad (2.15)$$

where γ denotes the gyromagnetic ratio (see Figure 2.5). In principle, only nuclei that possess a magnetic dipole moment μ (a non-zero spin) are relevant for MRI. As the hydrogen nucleus ^1H has the largest μ and a high natural abundance in biological systems, it is the most commonly used nucleus in MRI.

2.3 Probing diffusion with *magnetic resonance imaging* (MRI)

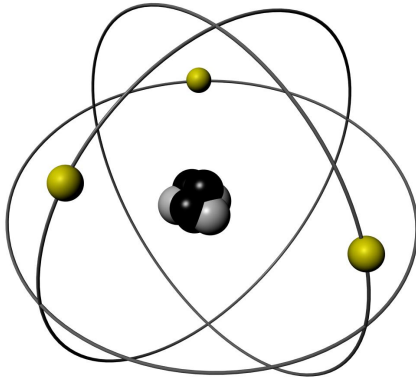


Figure 2.4: Bohr's model of an atomic nucleus with protons (black) and neutrons (grey) and its orbiting electrons (yellow) (taken from Ros et al. (2007)).

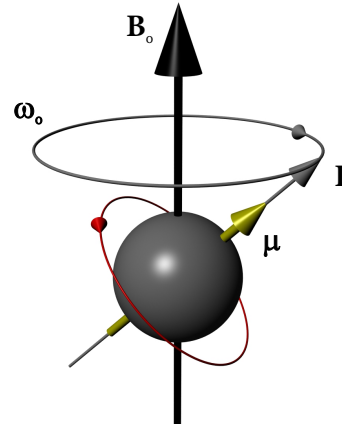


Figure 2.5: An atomic nucleus that possesses a nuclear spin I and the magnetic dipole moment μ precesses with the *Larmor frequency* ω_0 around the main magnetic field B_0 (adapted from Ros et al. (2007)).

By observing a large number of nuclei that are exposed to the main magnetic field B_0 , slightly more spins will align parallel than antiparallel to B_0 , as the parallel alignment is energetically more favorable. This gives rise to a longitudinal macroscopic magnetization M parallel to B_0 . In order to detect this macroscopic magnetization the thermodynamic equilibrium has to be disturbed by applying an additional *high frequency* (HF) magnetic field B_1 that is perpendicular to B_0 and oscillates with the *Larmor frequency* ω_0 . The application of the HF field B_1 tips the longitudinal magnetization M towards the transverse (x - y) plane perpendicular to B_0 (see Figure 2.6).

When the HF field B_1 ceases, tissue-dependent T_1 and T_2 relaxation processes take place simultaneously, the tipped magnetization M will return into equilibrium and a detectable NMR signal is induced in the receiver coils. The T_1 relaxation (spin-lattice relaxation) is governed by the interactions between spins and the lattice and describes the process in which the longitudinal magnetization regains equilibrium. In contrast, the T_2 relaxation (the spin-spin relaxation) results from interactions between the spins themselves, which gives rise to the decay of the transverse magnetization.

For MR imaging, the location of the signal has to be encoded in all three spatial directions x , y and z . This is achieved by applying additional, linear magnetic field gradients that superimpose the main magnetic field B_0 and introduce a spatially dependent variation of the magnetic field. The spatial encoding yields the frequency spectrum of the spin density, which can be

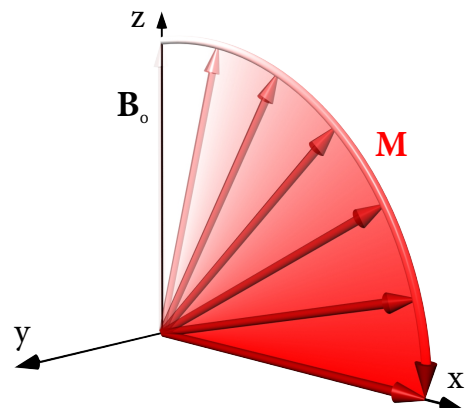


Figure 2.6: The application of an HF pulse tips the magnetization into the transverse plane perpendicular to B_0 .

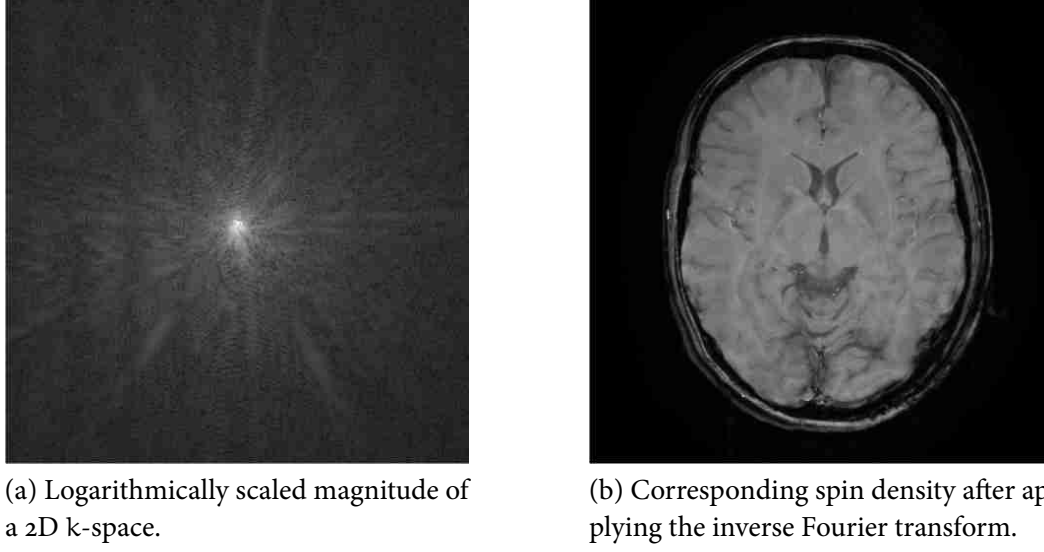


Figure 2.7: Example of a 2D k-space (a) and the corresponding transverse image of the spin density (b) (taken from Ros et al. (2007)).

described with the k-space formalism (Twieg, 1983). The k-space is a two-dimensional frequency space that is sampled successively according to a predetermined sampling path – the k-space trajectory. The consecutive sampling of k-space at discrete time points transforms the continuous MR signal into a discrete signal, thereby determining the spatial resolution of the final MR image that is reconstructed by applying an inverse Fourier transform to the fully sampled k-space. An example of a 2D k-space and the corresponding image of the spin density is given in Figure 2.7.

2.3.2 Diffusion weighted imaging (DWI)

Diffusion weighted imaging (DWI) is an MR imaging technique that measures diffusion and determines the mean-square displacement of diffusing molecules. In order to quantify diffusion, Stejskal and Tanner (1965) presented the so called *pulsed gradient spin echo* (PGSE) experiment, which is briefly illustrated in Figure 2.8. For this experiment, a 90° HF excitation pulse is applied to excite the spins and to tip the magnetization into the transverse plane. The signal is encoded by two diffusion sensitizing gradients that are symmetrically placed around a 180° refocusing HF pulse which is applied after time period τ starting from the excitation pulse. After another period of duration τ the signal is fully formed and can be acquired. The gradients are separated by the time interval Δ .

The application of a spatially varying, linear field gradient \mathbf{g} introduces a spatially dependent modulation of the precession frequency for spins in the main magnetic field \mathbf{B}_0 . As a result, the *Larmor frequency* ω_0 (see Eq. 2.15) becomes a function of position \mathbf{r} (Price, 1997):

$$\omega(\mathbf{r}) = \omega_0 + \gamma \mathbf{g} \mathbf{r} = \gamma \mathbf{B}_0 + \gamma \mathbf{g} \mathbf{r}. \quad (2.16)$$

2.3 Probing diffusion with *magnetic resonance imaging* (MRI)

In the general case, gradient \mathbf{g} is applied for a certain amount of time and causes the spins to dephase at different rates. This produces a cumulative phase shift (Price, 1997)

$$\varphi(t) = \gamma \int_0^t \mathbf{g}(t') \mathbf{r}(t') dt'. \quad (2.17)$$

Even though the gradient amplitude $\mathbf{g}(t)$ is in general time-dependent, it is, for simplicity, further assumed that the amplitude $\mathbf{g}(t)$ is constant over time: $\mathbf{g} = \mathbf{g}(t)$.

Going back to the PGSE experiment (see Figure 2.8), the first gradient \mathbf{g} , applied for the time duration δ , produces the phase shift (Price, 1997)

$$\varphi_1(\tau) = \gamma g \int_{t_1}^{t_1+\delta} \mathbf{r}(t) dt. \quad (2.18)$$

The following 180° HF pulse inverts the magnetization, while the second gradient, applied with identical gradient strength, rephases the spins and produces the phase shift

$$\varphi_2(\tau) = -\gamma g \int_{t_1+\Delta}^{t_1+\Delta+\delta} \mathbf{r}(t) dt. \quad (2.19)$$

Following Price (1997), the total phase shift is then given by

$$\begin{aligned} \varphi(2\tau) &= \varphi_1(\tau) + \varphi_2(\tau) \\ &= \gamma g \left(\int_{t_1}^{t_1+\delta} \mathbf{r}(t) dt - \int_{t_1+\Delta}^{t_1+\Delta+\delta} \mathbf{r}(t') dt' \right). \end{aligned} \quad (2.20)$$

In the absence of translational diffusion along the gradient direction, the cumulative phase shift has been completely reversed, the rephasing is complete and the total phase shift $\varphi(2\tau)$ is zero. However, if molecules have diffused along the direction of gradient \mathbf{g} during time period Δ , the rephasing will be incomplete ($\varphi(2\tau) \neq 0$) and signal attenuation will be observed.

Assuming the diffusion is Gaussian, the signal attenuation at position \mathbf{r} is given by the Stejskal-Tanner equation (Stejskal and Tanner, 1965)

$$S(\mathbf{r}) = S_0(\mathbf{r}) \exp(-bD(\mathbf{r})), \quad (2.21)$$

where S_0 is the signal without diffusion weighting ($b = 0$) and b , a diffusion weighting factor as introduced by Le Bihan and Breton (1985)

$$b = \gamma^2 \delta^2 g^2 \left(\Delta - \frac{\delta}{3} \right). \quad (2.22)$$

Hence, the measured diffusion signal depends on the diffusion coefficient D , the diffusion time Δ as well as on the amplitude and duration δ of the diffusion sensitizing gradient.

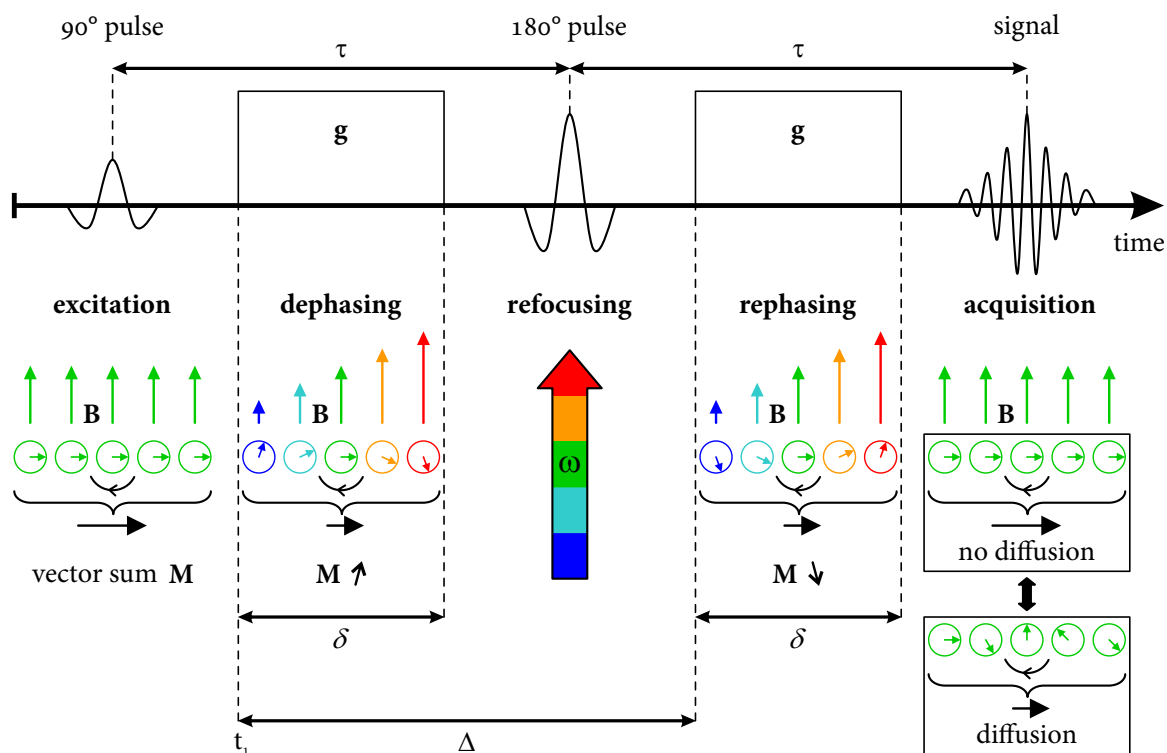


Figure 2.8: Schematic illustration of the diffusion weighted PGSE experiment (modified from Leemans (2006)). The strength of the magnetic field B is indicated by the length of the colored vertical arrows. During application of gradients g for time duration δ , the strength of the main magnetic field is modulated and the precession frequency ω becomes spatially dependent. After excitation with a 90° HF pulse and the application of the first gradient during time period τ , spins begin to dephase and precess at different frequencies ω due to the underlying, spatially varying magnetic field. The resulting spatially dependent dephasing of spins (illustrated with colored, encircled arrows) results in a loss of phase coherence and the net magnetization M decreases. The application of the 180° refocusing HF pulse in combination with the second gradient g during the second time period τ rephases the spins. After completion of both τ periods, the signal has finally formed and can be measured. In the absence of diffusion during the time Δ , no signal attenuation is observed and the cumulative phase shift has been completely reversed. However, if spins have undergone translational motion along the gradient direction during the onset of the gradients, rephasing of the spins will be incomplete and signal attenuation will be finally observed.

2.3 Probing diffusion with *magnetic resonance imaging* (MRI)

By rearranging Eq. 2.21, the diffusion coefficient $D(\mathbf{r})$ can be estimated from the diffusion weighted signal $S(\mathbf{r})$ and the unweighted signal $S_o(\mathbf{r})$

$$D(\mathbf{r}) = -b^{-1} \ln \left(\frac{S(\mathbf{r})}{S_o(\mathbf{r})} \right). \quad (2.23)$$

The computation of the diffusion coefficient with the Stejskal-Tanner equation (Eqs. 2.21–2.23) assumes unrestricted diffusion that can be modeled with a Gaussian distribution (see section 2.1). At a microscopic scale, however, diffusion in tissue is highly heterogeneous due to (micro-) cellular compartments and natural barriers (e.g., molecules, membranes, white matter fibers tracts). The observed diffusion coefficient is thus not solely guided by self-diffusion, but influenced by additional microscopic particles and structures. Consequently, the *apparent diffusion coefficient* (ADC) depends not only on the underlying diffusion coefficient of the tissue, but also on experimental parameters, such as diffusion time and voxel size (Le Bihan et al., 2006). If the diffusion time is sufficiently long, diffusion in tissue will be restricted and the ADC will be reduced.

2.3.3 *Diffusion tensor imaging* (DTI)

If the underlying tissue microstructure is isotropic, the diffusion coefficient $D(\mathbf{r})$ (Eq. 2.23) describes the diffusion properly. However, if the tissue structure is more complex and diffusion anisotropic, a scalar value is insufficient to characterize the diffusion and more complex models have to be used instead. In the mid-1990s, *diffusion tensor imaging* (DTI) was introduced by Basser et al. (1994), who generalized the Stejskal-Tanner equation (Eq. 2.21) to model diffusion by a second-order tensor (see section 2.1). The diffusion tensor $\underline{\mathbf{D}}(\mathbf{r})$ at position \mathbf{r} (compare Eq. 2.10) can be estimated by sampling the diffusion signal $S_k(\mathbf{r})$ in six or more linearly independent gradient directions \mathbf{g}_k with $k = 1, \dots, N$ along with the unweighted signal $S_o(\mathbf{r})$. The signal $S_k(\mathbf{r})$ for the k^{th} gradient direction is given by (Basser et al., 1994; Leemans, 2006)

$$S_k(\mathbf{r}) = S_o(\mathbf{r}) \exp \left(-b_k \hat{\mathbf{g}}_k^T \underline{\mathbf{D}}(\mathbf{r}) \hat{\mathbf{g}}_k \right), \quad (2.24)$$

where $\hat{\mathbf{g}}_k$ is the normalized gradient orientation vector

$$\hat{\mathbf{g}}_k = \frac{\mathbf{g}_k}{g} \quad (2.25)$$

and b_k the b value for the k^{th} gradient direction

$$b_k = \gamma^2 \delta^2 g_k^2 \left(\Delta - \frac{\delta}{3} \right). \quad (2.26)$$

Under the assumption that the gradient strength is identical for all gradient directions, b becomes independent from the gradient orientation

$$g = g_k \forall k \Rightarrow b = b_k \forall k. \quad (2.27)$$

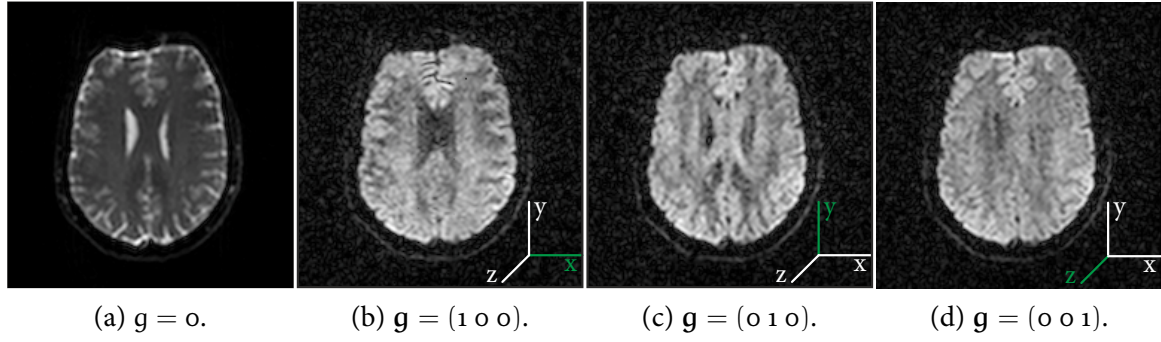


Figure 2.9: Transverse MRI images without diffusion weighting (a) and with diffusion weighting applied along the three axes with x = left-right, y = anterior-posterior and z = inferior-superior (b)–(d).

In Figure 2.9, images with and without diffusion weighting are presented for a human brain dataset.

The signals $S_k(\mathbf{r})$ (Eq. 2.24) for all $k = 1 \dots N$ gradient directions can be arranged as a linear system of equations and solved to estimate the apparent diffusions tensor $\underline{\mathbf{D}}(\mathbf{r})$. Decomposing the diffusion tensor yields quantitative, scalar parameters that represent distinct features of the tensor. Changes of single or multiple of these quantities are often an indicator for structural alterations in the brain, (e.g., reorganization processes, demyelination, axonal damage or changes in axonal density). Even though a variety of parameters have been proposed to characterize the diffusion tensor $\underline{\mathbf{D}}$ with a single value, only the most common, rotationally invariant measures as defined by Basser (1995); Pierpaoli and Basser (1996) and Bahn (1999) are presented:

- The *trace* is defined as the sum of the diagonal elements of $\underline{\mathbf{D}}$

$$trace = D_{xx} + D_{yy} + D_{zz} = \lambda_1 + \lambda_2 + \lambda_3. \quad (2.28)$$

- The *mean diffusivity*, also known as ADC_{mean} or simply ADC is the average diffusivity along all three axes:

$$ADC = \langle \lambda \rangle = \frac{Trace}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}. \quad (2.29)$$

- The *fractional anisotropy* (FA) is a measure that captures the degree of diffusion anisotropy

$$FA = \sqrt{\frac{3}{2} \frac{\sqrt{(\lambda_1 - \langle \lambda \rangle)^2 + (\lambda_2 - \langle \lambda \rangle)^2 + (\lambda_3 - \langle \lambda \rangle)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}}. \quad (2.30)$$

The FA takes values between 0 and 1, whilst 0 corresponds to isotropic and 1 to anisotropic diffusion. In order to visualize the strength and directionality of the dif-

fusion, the FA can be color-coded with the RGB color-model⁴. For this purpose, the principal direction of diffusion (v_1) is interpreted as RGB value and weighted with the FA. The color is then an indication for the principal orientation of diffusion (red = left-right, green = anterior-posterior and blue = inferior-superior), whilst color intensity determines the strength of diffusion.

- The *axial diffusivity* (AD) is the diffusivity along the principal axis

$$AD = \lambda_1. \quad (2.31)$$

- The *radial diffusivity* (RD) describes the diffusivity perpendicular to the principal axis of diffusion:

$$RD = \frac{\lambda_2 + \lambda_3}{2}. \quad (2.32)$$

An example for diffusion tensors and the presented tensor-based quantities is presented in Figure 2.10 for a single slice of an *in vivo* human brain dataset.

However, the measured diffusion signal is just an approximation of the underlying microstructure that consists of billion axonal connections per imaging voxel (see section 2.2). Even though the diffusion tensor is a good approximation when the tissue is homogeneous and axons are aligned parallel, it is insufficient in heterogeneous tissue, where multiple fiber compartments give rise to a complex diffusion signal (e.g., crossing, fanning or kissing fibers). In order to describe the diffusion in the presence of complex fiber architectures accurately more elaborate methods have to be used. With complex techniques that sample the diffusion signal in many directions with high b-values (e.g., *high angular resolution diffusion imaging* (Tuch, 2004; Descoteaux et al., 2007)), *orientation distribution functions* (ODFs) can be reconstructed that characterize the diffusion signal more accurately and are able to resolve regions with multiple fiber compartments.

2.4 Tractography

Fiber tractography, also referred to as fiber tracking, is an MRI post-processing technique for the non-invasive reconstruction of the white matter fiber network. The MRI diffusion signal is utilized to trace the path of diffusion and to determine fiber trajectories in three-dimensional space. Figure 2.11 shows an illustration for a single fiber tract that traverses along the principal direction of diffusion.

In comparison to voxel-based MR imaging modalities, fiber tracking generates a three-dimensional representation of the underlying microstructure that models the microscopic axonal connections in the brain. This enables not only a superior visualization of the fiber network, but offers various novel ways to study the white matter neuroarchitecture *in vivo*. While fiber tractography, can be used to reconstruct and investigate the fiber network of the

⁴RGB is an additive color-model in which varying color components of red, green and blue are added to produce distinct colors.

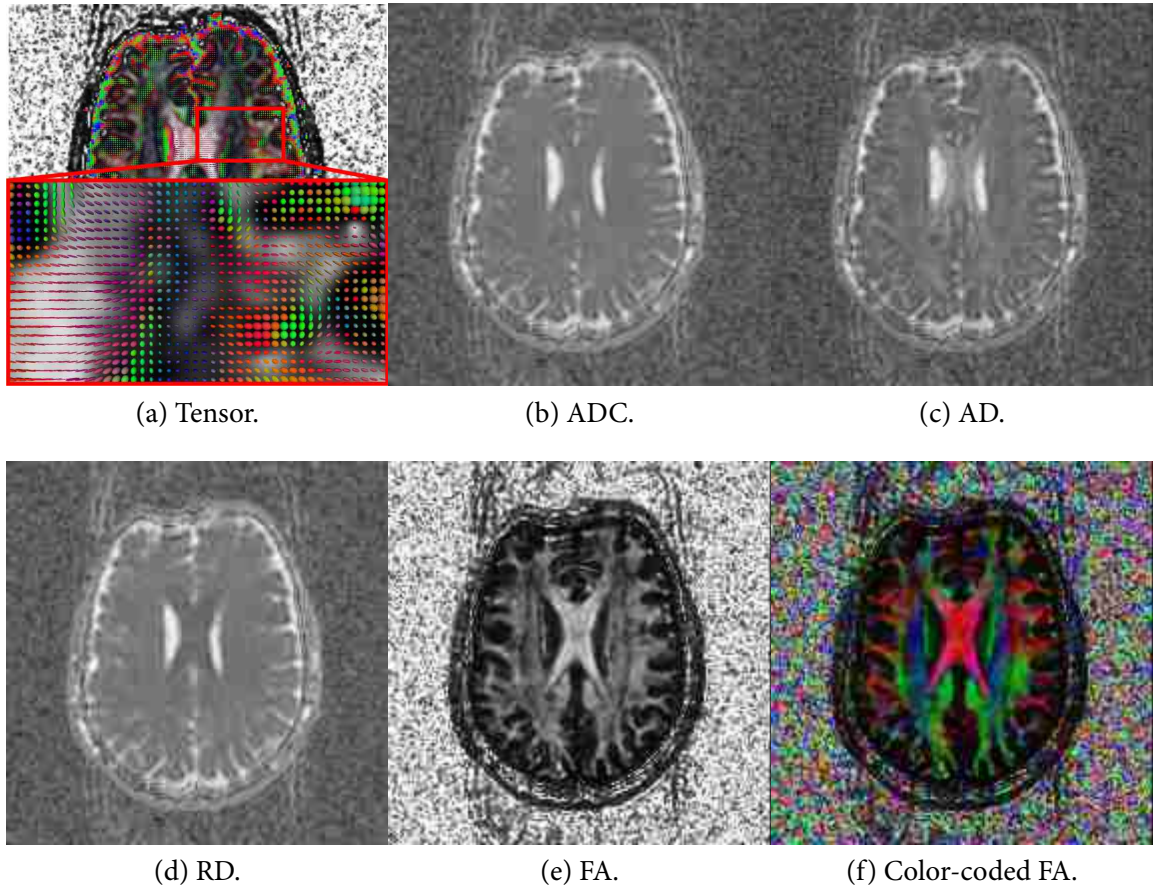


Figure 2.10: Example for a diffusion tensor image (a) and various contrasts derived from diffusion tensors (b)–(f). The diffusion ellipsoids (a) and the color-coded FA (f) are colored according to the principal direction of diffusion with red = left-right, green = anterior-posterior and blue = inferior-superior.

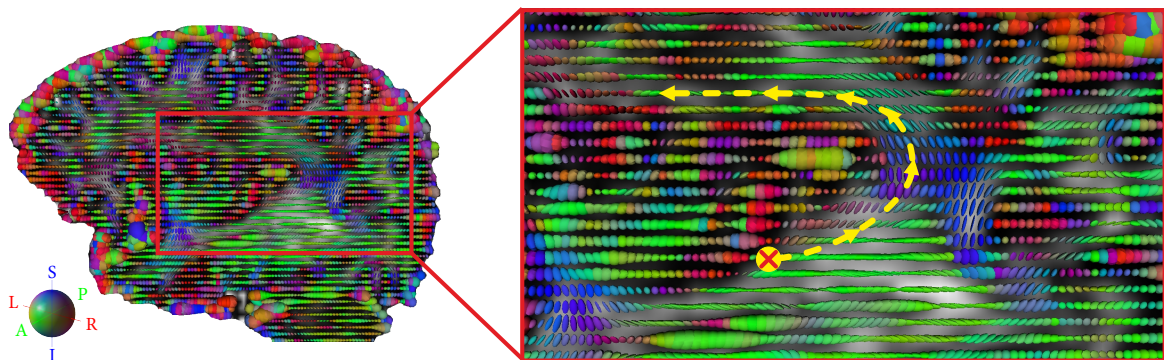


Figure 2.11: Illustration of tensor based fiber tractography for one tract. By starting at the seed point (marked with a red \times), the trajectory of the tract is propagated along the principal direction of diffusion.

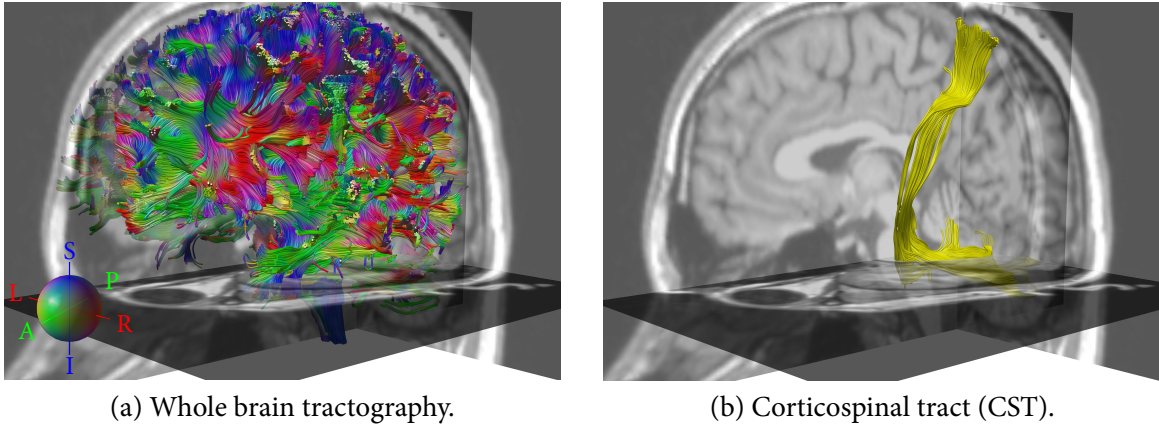


Figure 2.12: Example for whole brain tractography (a). Each tract in the dataset is colored to encode its orientation. An example for a single fiber bundle, depicted in yellow, is shown in (b).

entire brain, it also permits targeted tractography in which only selected white matter fiber bundles are extracted (for an example see Figure 2.12). As pointed out by Jones and Pierpaoli (2005), tractography can visualize both the structure and the diffusion properties of white matter pathways simultaneously. It can not only be used to investigate the structural integrity of white matter or to perform preoperative, neurosurgical planning (Clark and Byrnes, 2011), but might also be helpful to delineate lesions or tumors in the white matter. Some studies already showed that tractography is beneficial to assess the structural architecture of white matter pathways with tumor involvement (Stadlbauer et al., 2007; Nucifora et al., 2007).

Tractography can be performed with a huge variety of tracking algorithms that initiate the fiber tracking by placing single or multiple seed points in the voxels of the dataset. When the entire brain is specified as seed region, this is usually referred to as whole brain tractography. From the seed regions, the fiber tracking is performed voxel-by-voxel until the tracts leave the tissue, regions become too unreliable (e.g., low FA) or alternative stopping criteria are satisfied. For the fiber tracking, either deterministic or probabilistic approaches can be used. While deterministic methods propagate the trajectory of the tract according to the best estimation of the fiber orientation in the underlying microstructure until the tract terminates (Alexander, 2011), probabilistic approaches take sources of uncertainty for the estimation of the diffusion orientation into account and utilize probabilistic models and iterative procedures (e.g., *Markov chain Monte Carlo*, bootstrapping) for the tracking (Parker, 2011).

Irrespective of the tracking technique, the result is a set of fiber tracts, denoted by \mathbb{D} , that have no apparent structural organization and are not directly associated to each other. Each single tract $p \in \mathbb{D}$ is represented by a set of N tract points $p = \{p_1, \dots, p_N\}$ that define

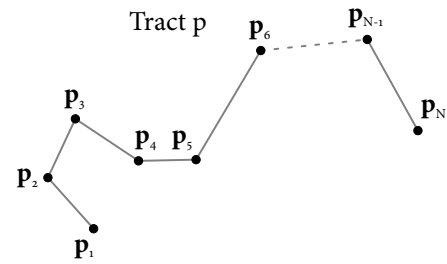


Figure 2.13: Illustration and formal definition of a fiber tract p with N tract points. The tract points that define the tracts are depicted in black while the individual segments of the tract are shown in grey.

p's trajectory (see Figure 2.13 for an illustration). For completeness, a fiber bundle is defined as a collection of fiber tracts that follow a similar anatomical path.

While diffusion MRI provides the unique opportunity to measure diffusion *in vivo*, the technique is inherently impaired by a low spatial resolution. Due to physical limitations (e.g., inevitably long diffusion times, signal decay) as well as the high number of acquisitions that are necessary to accurately assess anisotropic diffusion, acquisition time is prolonged and higher resolutions are prevented. As tractography is based on the relatively low resolution diffusion signal⁵, it is confronted with partial volume effects and data imperfections that may cause incomplete, disrupted or wrongly connected fiber bundles.

Reconstructed fiber tracts are only an approximation of the underlying microscopic axonal connections. The verification of the correctness of reconstructed fibers is thereby a major problem, because a gold standard cannot be obtained in human subjects. Even though Lawes et al. (2008) showed a good correspondence between various bundles from tractography and *post-mortem* human brain specimen using blunt dissection, Lawes and Clark (2011) also emphasized that the bundles under study correspond to regions with a simple anatomy in which axons run mainly parallel. In white matter regions with a complex underlying microstructure blunt dissection is prone to errors. Lawes and Clark (2011) also pointed out that different tracts will be different in the same species, while same tracts will differ among different species. Results for animal or even phantom studies are therefore of limited use to prove the correctness of reconstructed fiber bundles in humans.

2.5 Quantitative analysis of diffusion MRI data

Diffusion weighted MRI is an important tool to characterize and study a variety of disorders that are associated with alterations in brain diffusivity. It is not only relevant for a prompt diagnosis of acutely occurring disorders (e.g., acute stroke, traumatic brain injuries) but also for diagnosis, treatment planning and clinical monitoring of slowly manifesting diseases. In fact, the range of applications for DWI is widespread and not limited to the investigation of pathological conditions (e.g., psychiatric disorders, neurodegenerative diseases). It is also used to study natural changes, for example during brain development (Löbel, 2007) and white matter maturation (Moseley, 2002).

In order to detect subtle alterations in the diffusivity, most studies perform a statistical analysis for two groups of subjects by relying on quantitative measures that are directly derived from the diffusion tensor (e.g., FA, ADC, RD). Both, the experimental and the control group consist of a set of subjects, who are selected in compliance with certain criteria (e.g., age, gender, handedness). The statistical group analysis of the diffusion MRI data is then usually conducted with one of four major approaches:

- *region of interest* (ROI)-based methods,

⁵Nevertheless, it should be noted that a voxel with 1 mm³ voxel size still contains approximately 1 billion axonal connections.

- *voxel-based morphometry* (VBM)-styled analysis techniques (Ashburner and Friston, 2000),
- *tract-based spatial statistics* (TBSS) (Smith et al., 2006) or
- *fiber bundle-driven techniques* (FDTs).

The most basic, but still very popular methods for group analysis are ROI-based techniques, where the statistical analysis is performed in manually or semi-automatically delineated *regions of interest*. The quantitative values are extracted from the ROIs of each subject, before the statistical analysis is performed between the two groups. ROI-based methods offer the advantage that they are widely applicable, irrespective of a disease. As long as the ROIs can be adequately placed, the analysis can be conducted, even if the *regions of interest* are dislocated, for example due to space occupying lesions. However, since this ROI placement is often performed by hand and individually for each subject, the procedure is highly time consuming and an operator with profound anatomical knowledge is mandatory to place the ROIs correctly. The placement of the ROIs is indeed the most critical part of the analysis and incorrect placement may result in significant errors. In fact, studies showed that the accuracy and the outcome of the statistical analysis are not only influenced by the underlying differences between the groups in certain regions but also by the placement of ROIs (Kanaan et al., 2006; Ozturk et al., 2008). Depending on the investigated region and the experience of the operator, slight to substantial differences between ROIs and the extracted values are observed, even if the operator is supposed to repeatedly place ROIs in the same regions of the same dataset (Ozturk et al., 2008).

In order to superimpose corresponding brain structures of different datasets and account for inter-subject variability of brain anatomy, the majority of the other approaches (VBM, TBSS and FDT) rely on image registration that spatially normalizes the datasets of different subjects. While both linear and nonlinear techniques can be utilized to transform the datasets, the application of nonlinear normalization techniques enables a more accurate superposition of corresponding brain structures. Even though various coregistration techniques achieve good normalization results and high spatial agreement in healthy brains (Klein et al., 2009), they often have difficulties in handling space-occupying lesions in non-healthy brains. In cases in which the normalization fails, ROI-based methods have to be used instead. Performing coregistration causes not only spatial normalization of the brain datasets, but also the interpolation of the quantitative values that are subject to the analysis. This effect is measurable and more pronounced at the boundaries of brain structures such as grey and white matter or at the boundaries of white matter bundles (Chao et al., 2009).

Voxel-based morphometry (Ashburner and Friston, 2000) is a fully automated, computational approach to perform the quantitative analysis voxel-by-voxel for the entire brain. The individual datasets are spatially normalized and transferred into a common template space. The tissue is segmented into *cerebrospinal fluid*, grey and white matter and the data is spatially smoothed in order to ensure the validity of the subsequent statistical analysis, which is supported by *random field theory* to correct for multiple comparisons. Even though VBM enables fast and user-independent whole brain processing, it is sensitive to misregistration and segmentation artifacts, which may result in statistically significant differences at tissue

boundaries. While spatial smoothing is necessary to ensure that the statistical analysis is valid for VBM, it increases the partial volume effect. The spatial smoothing is also a possible source of error, because different smoothing parameters can greatly affect the results (Jones et al., 2005). Since there was a controversial debate on VBM and its advantages and disadvantages, the interested reader is directed to the dialog between Bookstein (2001) and Ashburner and Friston (2001).

TBSS was developed by Smith et al. (2006) to solve the problems associated with VBM. Like VBM, TBSS is also fully automated and performs a spatial normalization at the beginning. However, in order to circumvent misregistration or segmentation issues, TBSS performs the quantitative analysis solely on a white matter skeleton that is extracted from the average FA maps of all subjects. The statistical analysis of TBSS is notably more robust compared to VBM, but is also restricted to the analysis of white matter structures. Moreover, the analysis is limited to the FA skeleton, which is only one pixel wide and represents only the most prominent part of a fiber bundle. Thus, alterations in the periphery of a bundle cannot be detected with TBSS.

A different approach is based on the idea to incorporate fiber tracts into the quantitative analysis (Fillard et al., 2003; Gong et al., 2005; O'Donnell et al., 2007). Instead of performing the analysis for the whole brain or the white matter, it is conducted selectively for individual fiber bundles. Tractography is hereby primarily used to identify voxels that belong to the *fiber bundle of interest*. Then, a mean tract is computed for the fiber bundle and the quantitative values that belong to the fiber bundle are projected onto this mean tract. The statistical analysis between multiple subjects is finally performed along this mean tract, which results in a one-dimensional representation of the differences along the trajectory of the bundle. Even though the results of this type of processing are easy to understand, fiber-based methods have severe limitations. Due to the use of a mean tract projection and the resulting reduction to one dimension, the method is limited to the analysis of fiber bundles with a simple tubular structure. Another important issue concerns the selection of the fiber bundles, which is often difficult to automate and usually performed manually. In this case, shape and appearance of the bundle are influenced by the user, which again is prone to errors (Heiervang et al., 2006) and might lead to bundles that are not necessarily comparable. If bundles are incomplete or the trajectories deviate from the trajectories of other bundles, this has to be taken into account. It might even be necessary to exclude such bundles from the analysis.

While most *fiber bundle-driven techniques* use spatial normalization to align datasets of different subjects, methods exist that perform the quantitative analysis in the native space of the datasets. In this case, setting up the correspondence between the bundles of different subjects is a major problem. Since fiber bundles vary in shape, size, extent and curvature among subjects, the spatial correspondence has to be set up manually, which is not a trivial but a cumbersome task (Gong et al., 2005).

2.6 Cluster analysis and similarity measures

Cluster analysis encompasses a set of unsupervised machine learning techniques that partition objects into groups according to the similarity of object specific properties. The roots of cluster analysis can be traced back as far as to the ancient Greeks when Aristotle attempted to create

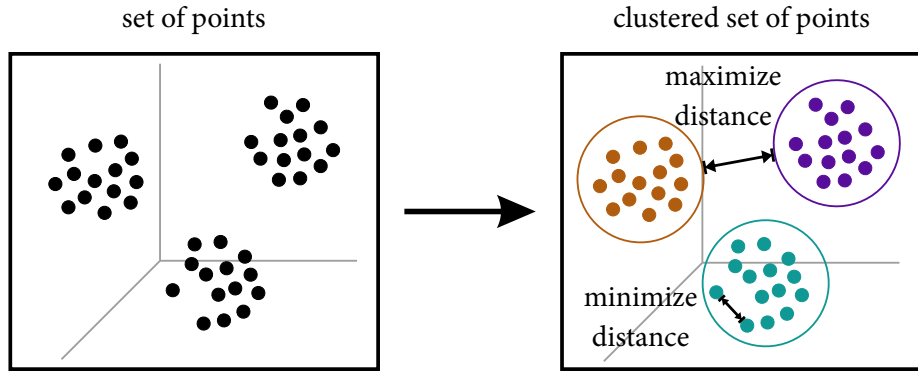


Figure 2.14: Fundamental concept of cluster analysis. The principal goal of cluster analysis is to group a set of objects (e.g., a set points) into clusters according to their similarity. During the iterative clustering process, distances between clusters are maximized, while differences between objects of the same cluster are minimized.

a taxonomy of living objects (Andreopoulos et al., 2009). Later in the 18th century, cluster analysis was intensively studied by Buffon, Cuvier and Linné (Hansen and Jaumard, 1997), but it was not before 1960 when an explosion of interest in cluster analysis was noticed (Blashfield and Aldenderfer, 1978). With the increased availability of modern general purpose computing systems and capable software packages at the beginning of the 21st century, this interest continued to increase and led to a rapid growth in publications (Xu and Wunsch, 2008). As a result, cluster analysis has been widely used in various disciplines as a tool for exploratory data analysis (e.g., in natural sciences, psychology, medicine, engineering) (Hansen and Jaumard, 1997).

The aim of cluster analysis is to group elements of a dataset into distinct clusters or groups in which differences between clusters are maximized (high inter-cluster heterogeneity) and differences between elements of a single cluster are minimized (high intra-cluster homogeneity) (see Figure 2.14). For this kind of partitioning, clustering techniques employ similarity measures that determine the similarity between objects by comparing specific and distinguishable properties or features of these objects. The employed features are generally unique properties of the objects (e.g., length, weight). On the basis of existing features, additional and meaningful features can be generated using feature extraction (Xu and Wunsch, 2008). As illustrated in Figure 2.15, the relationship between features and similarity measures can be formally expressed with the notation of the *unified modeling language* – UML (Object Management Group, 2011).

Formally, a function that determines the similarity of two objects p and q from a dataset \mathbb{D} on the basis of features, is called a distance or similarity function $d(p, q)$ if the function is symmetric, positive semidefinite and reflexive:

$$\text{symmetric: } d(p, q) = d(q, p), \quad (2.33)$$

$$\text{positive semidefinite: } d(p, q) \geq 0 \quad \forall p, q \in X, \quad (2.34)$$

$$\text{reflexive: } d(p, q) = 0 \quad \text{if } p = q. \quad (2.35)$$

(feature extraction)

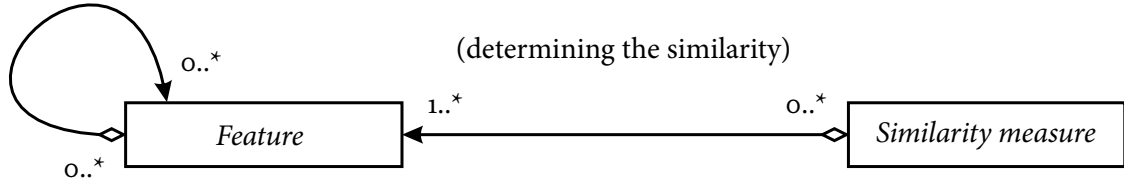


Figure 2.15: Depiction of the relationship between features and similarity measures using UML. Similarity measures utilize features to determine the similarity between objects. On basis of existing features, feature extraction generates additional, meaningful features.

From Eqs. 2.34 and 2.35, it implicitly follows that the distance between two objects decreases the more similar they are, while it increases in the opposed case. By convention, a distance function that adheres to Eq. 2.33–2.35 is called a semimetric. If the triangle inequality

$$d(p, q) \leq d(p, r) + d(r, q) \quad \forall p, q, r \in X \quad (2.36)$$

is also satisfied, it is named a metric (Xu and Wunsch, 2008). Even though metrics constitute an important group of similarity measures, they are mostly utilized to assess the spatial distance between objects during the cluster analysis (e.g., distance between points in n -dimensional space). In contrast, non-metric measures are also often used but measure other relations than the spatial distance between objects⁶.

Over the years, a multitude of different cluster analysis algorithms have been developed that are suitable for various problem domains and different types of data. Methods were presented that rely on partitioning (e.g., *k-means* (Forgy, 1965)), graph theoretical approaches (e.g., CLICK (Sharan and Shamir, 2000)), fuzzy logic (e.g., *fuzzy c-means* (Bezdek, 1981)) or neural networks (e.g., *adaptive resonance theory* (Carpenter and Grossberg, 1988)) (see Xu and Wunsch (2008); Everitt et al. (2011)). However, this thesis primarily focuses on *hierarchical agglomerative clustering*. In *hierarchical agglomerative clustering*, a bottom-up approach is used to generate the clusters. Initially, every object in the dataset is considered an individual singular cluster with only one object. During the iterative clustering process, the two most similar clusters are merged together in each iteration until a predefined stopping criterion is satisfied. For *hierarchical agglomerative clustering*, the pairwise similarities between all objects of a dataset have to be determined. In practice, this is equivalent to the computation of the similarity matrix \mathcal{S} that holds the similarities between all elements of \mathbb{D} . The pairwise similarity between two objects x_i and x_j of \mathbb{D} (with $i, j = 1 \dots N$) is computed with the similarity function $d(x_i, x_j)$ and represented by a single value in matrix \mathcal{S} at the $(i, j)^{\text{th}}$ position. If \mathbb{D} consists of N elements, the similarity matrix has size $N \times N$ (compare Table 2.1). Due to the symmetry of d (Eq. 2.33) only the upper matrix has to be computed. From Eq. 2.35 it follows that the main diagonal elements $\mathcal{S}(i, i)$ for $i = 1 \dots N$ remain zero. Hence, only $(N^2 - N)/2$ matrix elements have to be computed to assess the similarity for all elements in \mathbb{D} .

⁶Various examples for non-metric similarity measures can be found in the publication by Skopal and Bustos (2011).

2.7 Parallel computing

2.7.1 General definitions

Traditionally, parallel computing can be defined as the concurrent use of multiple processing elements to perform computations simultaneously. To differentiate between sequential and parallel computer architectures Flynn (1972) introduced a general classification scheme, which is nowadays known as Flynn's taxonomy. Flynn defined four different classifications (compare Table 2.2) with respect to the concurrency of instruction and data in the computer architecture:

- *Single instruction stream over single data stream (SISD)*
Computer with SISD architecture are sequential computers with a single *central processing unit* (CPU) as the classical *Von Neumann architecture* (von Neumann, 1945). They do not exhibit any concurrency.
- *Multiple instruction streams over single data stream (MISD)*
MISD systems are architectures that consist of multiple instruction streams that operate on the same single data stream. MISD architectures are rarely used and merely play a role in redundant systems to achieve fault tolerance.
- *Single instruction streams over multiple data streams (SIMD)*
SIMD systems consist of multiple processing units that perform the same instruction on multiple data streams in parallel.
- *Multiple instruction streams over multiple data streams (MIMD)*
MIMD computers are multiprocessing systems that have multiple independent CPUs or multi-core CPUs. The processing cores can execute different instructions on different data streams. These architectures represent the most common architecture for personal computers nowadays.

Besides the architecture of a system, algorithms can be classified according to their implementation into sequential and parallel algorithms. A sequential algorithm performs a task strictly

Table 2.1: $N \times N$ similarity matrix $\$$ for a dataset \mathcal{D} with N objects. An element of the matrix at the $(i, j)^{\text{th}}$ position specifies the similarity between the two objects x_i and x_j computed with a similarity function d that adheres to the general definition of similarity measures (Eqs. 2.33–2.35).

	x_1	x_2	x_3	\dots	x_{N-1}	x_N
x_1	0	$d(x_1, x_2)$	$d(x_1, x_3)$	\dots	$d(x_1, x_{N-1})$	$d(x_1, x_N)$
x_2		0	$d(x_2, x_3)$	\dots	$d(x_2, x_{N-1})$	$d(x_2, x_N)$
x_3			0	\dots	$d(x_3, x_{N-1})$	$d(x_3, x_N)$
\vdots				\ddots	\vdots	\vdots
x_{N-1}					0	$d(x_{N-1}, x_N)$
x_N						0

Table 2.2: Flynn's taxonomy that differentiates sequential and parallel architectures according to the concurrency of instructions and data access.

	Single data	Multiple data
Single instruction	<i>Single instruction stream over single data stream (SISD)</i>	<i>Single instruction streams over multiple data streams (SIMD)</i>
Multiple instruction	<i>Multiple instruction streams over single data stream (MISD)</i>	<i>Multiple instruction streams over multiple data streams (MIMD)</i>

in sequential order, whilst a parallel algorithm employs concurrency to solve a separable task in parallel. Parallel algorithms are further divided into task and data parallel algorithms. While task parallel algorithms are based on the notion to parallelize different tasks and process the tasks independently, data parallelism adheres to the principle to process the data in parallel using the same instruction for each data element. Hence, task parallel algorithms are ideal for MIMD architectures while data parallel algorithms strongly benefit from the raw processing power and simplicity of SIMD architectures.

To analyze the performance of a parallel algorithm, the speedup can be computed. Given a parallel algorithm, the achievable speedup of the algorithm is determined only by the execution time T_s of the sequential algorithm and the execution time T_P of the parallel algorithm that employs P processing cores. It is then defined as:

$$S_P = T_s / T_P. \quad (2.37)$$

When the speedup S_P equals the number of processing cores P the speedup is called linear or ideal. If the speedup S_P is even higher than the number of processing cores P , it is called super linear speedup. The sequential program is often identical to the parallel program but uses only one processing core. However, the sequential program might be different if a more efficient implementation is available for the single core system.

2.7.2 Tasks, processes and threads

The resources of the system and tasks are usually controlled by the *operating system*. Modern multitasking *operating systems* utilize sophisticated scheduling schemes to perform multitasking and execute tasks simultaneously. Fast task context switches are initiated multiple times per second to simulate a quasi concurrent execution of the tasks, even if only one processor is installed in the system (SISD architecture). If more than one processor is available, true multitasking can be performed. At this point a task denotes either a process or a thread of execution. A process is the instance of a static computer program, whereas a thread of execution is a lightweight process and the smallest unit that can be executed by the scheduler. A thread is created by its parent process and can usually only exist as long as the parent process endures. Resources are shared between all threads of a process and intra-process communication between threads is possible. Since a thread is lightweight, the context switches between

threads of the same process are usually faster than context switches between processes. Thus, a process is multithreaded if more than one thread exists in the context of the process. If only threads of a single process are executed in parallel, this is called thread level parallelism.

2.7.3 GPU computing and CUDA

The *graphics processing unit* (GPU) is a highly specialized device for the manipulation of computer graphics. Tasks and algorithms that are intended for computer graphics are often performed by applying identical, atomic operations to each voxel. While all voxel-based operations are independent, they can easily be parallelized and manufacturer started to incorporate stream processors (SIMD architecture) into the GPUs to enhance the performance.

In 2006, the *compute unified device architecture* (CUDA) was presented by Nvidia (NVIDIA Corporation, Santa Clara, CA, USA) as a general purpose parallel computing architecture that incorporates highly parallelizable stream processors (NVIDIA, 2010). With the introduction of CUDA, Nvidia presented the new execution model SIMT (*single instruction stream, multiple threads*) and a C based framework for the software development with CUDA-enabled GPUs. As pointed out by NVIDIA (2010) the SIMT architecture is highly similar to the SIMD architecture and the additional SIMT behavior can be neglected for the purpose of correctness. For the sake of simplicity, the focus remains therefore on the SIMD architecture. A detailed differentiation between SIMD and SIMT can be found in the technical report by NVIDIA (2010).

CUDA-enabled GPUs consist of a set of multithreaded streaming multiprocessors where any multiprocessor can handle hundreds of threads concurrently (NVIDIA, 2010). Each stream processor has a single control unit that controls the computation of multiple processing cores that work in parallel (see Figure 2.16). When data is transferred to the CUDA device and the program is executed, the data is processed by all available multiprocessors.

SIMD architectures are usually well suited for data parallel problems or problems that can be vectorized. However, if a task contains branching or divergent execution paths, the performance degrades and MIMD architectures might be more suited. If a task cannot be parallelized, SISD or MIMD architectures are preferable since CPUs are generally faster than single threaded stream processors.

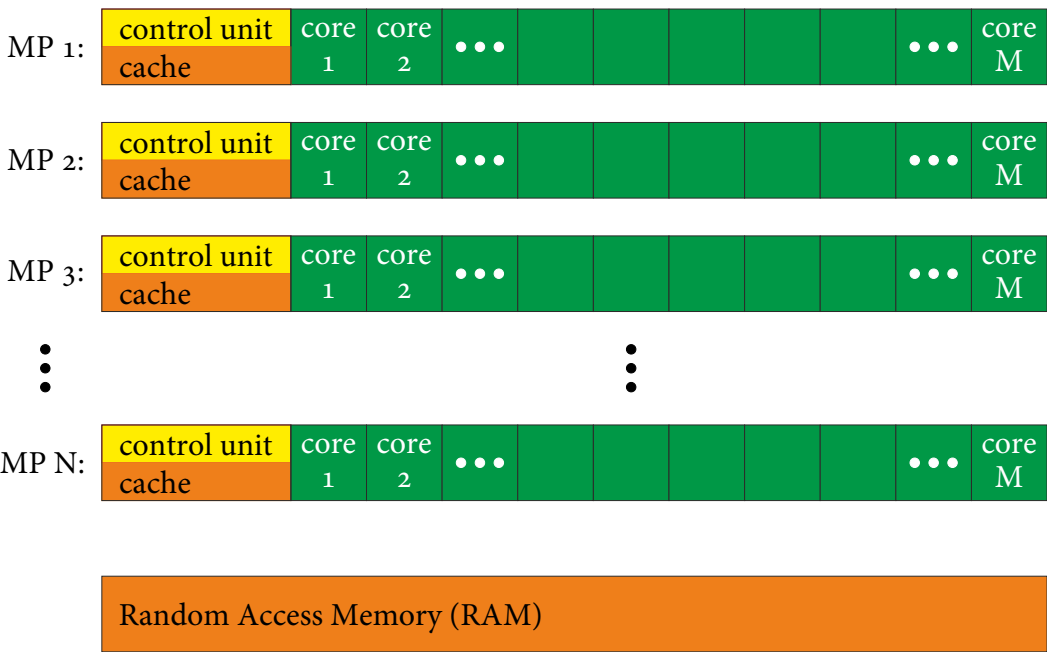


Figure 2.16: CUDA-enabled GPU architecture (adapted from NVIDIA (2010)). GPUs consist of a set of multithreaded streaming multiprocessors (MPs), each equipped with a single control unit that controls the parallel processing computation of multiple processing cores.

Part II

Cluster analysis of fiber tracts

3 Cluster analysis of large tractography datasets

*Divide each difficulty into as many parts
as is feasible and necessary to resolve it.*

— René Descartes (1596–1650)

Cluster analysis of fiber tracts is limited by the inherent complexity of fiber tracts and by the large number of tracts that have to be processed. In order to make cluster analysis applicable to large tractography datasets, this chapter presents a new clustering framework for fiber tracts. Concepts for the processing of large datasets are described and a method for the integration of anatomical information into the cluster analysis is presented.

In the previous chapter, the basics of cluster analysis and similarity measures have been reviewed for generic objects (see section 2.6). This thesis, however, primarily focuses on the cluster analysis of fiber tracts (see section 2.4). Beginning with this chapter, the term fiber tract will therefore be used instead of the more generic term object. Nevertheless, the presented methods are not restricted to fiber tracts and can be used for grouping all kind of objects (e.g., points, documents). In this chapter, the general definition of similarity functions, as introduced in section 2.6, is used. For the actual clustering of a dataset, explicit similarity functions that satisfy the general definition have to be used. A selection of explicit similarity measures is presented later in chapter 4.

3.1 Introduction

The application of clustering for automated bundling of neural pathways was pioneered by Shimony et al. (2002) and Ding et al. (2003). While first approaches by Shimony et al. (2002) were driven by the idea to use fuzzy clustering together with Euclidean distance and shape similarity measures, Ding et al. (2003) employed conventional *hierarchical agglomerative clustering* and a similarity measure that was based on corresponding tract segments.

One fundamental limitation of clustering is the high computational complexity, which is immanent to the majority of conventional clustering algorithms (Xu and Wunsch, 2008).

Compared to the classic cluster analysis problem, where the data often consists of single points in n -dimensional Euclidean space, fiber tracts are far more complex and consist of sets of points that constitute complex trajectories in 3D space (cmp. section 2.4 Tractography). Appropriate measures are therefore indispensable to determine the similarity between fibers. However, both the costly clustering and the necessarily more complex similarity measures increase the total computational load and typically restrict cluster analysis to small datasets that consist of only a few thousand fiber tracts.

In recent years, a multitude of methods have been proposed for both classification and clustering of fiber tracts (Ding et al., 2003; Brun et al., 2004; Moberts et al., 2005; Huang et al., 2005; O'Donnell et al., 2006; Zhang et al., 2008; Maddah et al., 2008; Li et al., 2010; Garyfallidis et al., 2012; Wu et al., 2012; Wang et al., 2013). The first clustering approaches solely relied on similarity measures to group tracts into bundles (e.g., Ding et al. (2003), Moberts et al. (2005)). Various researchers investigated spectral clustering approaches (Brun et al., 2004; O'Donnell et al., 2006) and used *spectral embedding* to map the fibers to three-dimensional Euclidean space, which resulted in more easy algorithmic handling of the inherent complexity of fiber tracts (O'Donnell et al., 2006). These first methods primarily focused on clustering of single subject data and neglected anatomical information. Later on, researchers started to experiment with the clustering of multiple input datasets and the incorporation of anatomical features into the clustering. While O'Donnell and Westin (2007), for example, performed multi-subject clustering to create an atlas that was used to automatically label fiber tracts, Maddah (2008) developed an expectation-maximization algorithm and used Bayesian modeling to integrate spatial anatomical information. More recent approaches focused on repeated, simultaneous clustering of multiple datasets (Visser et al., 2011) and fast voxel clustering of rasterized tracts (Guevara et al., 2011), but neglected anatomical correspondence of fibers and obtained clusters. Overall, despite the multitude of available methods that have been proposed for both classification and clustering of fiber tracts, fast, consistent and anatomically correct clustering for multiple subjects is still challenging.

To overcome these shortcomings, a new clustering framework is presented that introduces the novel cluster analysis technique CATSER (*cluster analysis through smartly extracted representatives*). While conventional clustering techniques are often limited by long processing times, CATSER is characterized by low computational complexity and is applicable to large tractography datasets that contain hundreds of thousands of fiber tracts. In order to reduce the computation time, the approach relies on random sampling, partitioning of the data and parallel computing.

In practice, clustering of tracts is rarely optimal. As the outcome of cluster analysis is influenced by various factors, such as the similarity measure, the clustering parameters and the data itself, it is challenging to set up the analysis in a way that consistent and reproducible clustering of different datasets with a good correspondence to anatomical fiber bundles is achieved. Due to the high variability of different datasets, fiber bundles are often divided into various parts, or unrelated bundles are falsely merged. It is in fact unlikely that clustering without anatomical guidance can be used to extract fiber bundles reliably so that the generated bundles are anatomically correct for all datasets. Hence, consistent, reproducible and correct extraction of fiber bundles across multiple subjects, which is solely based on tract similarity and utilizes no anatomical guidance, is difficult to achieve.

3.2 Cluster analysis through smartly extracted representatives (CATSER)

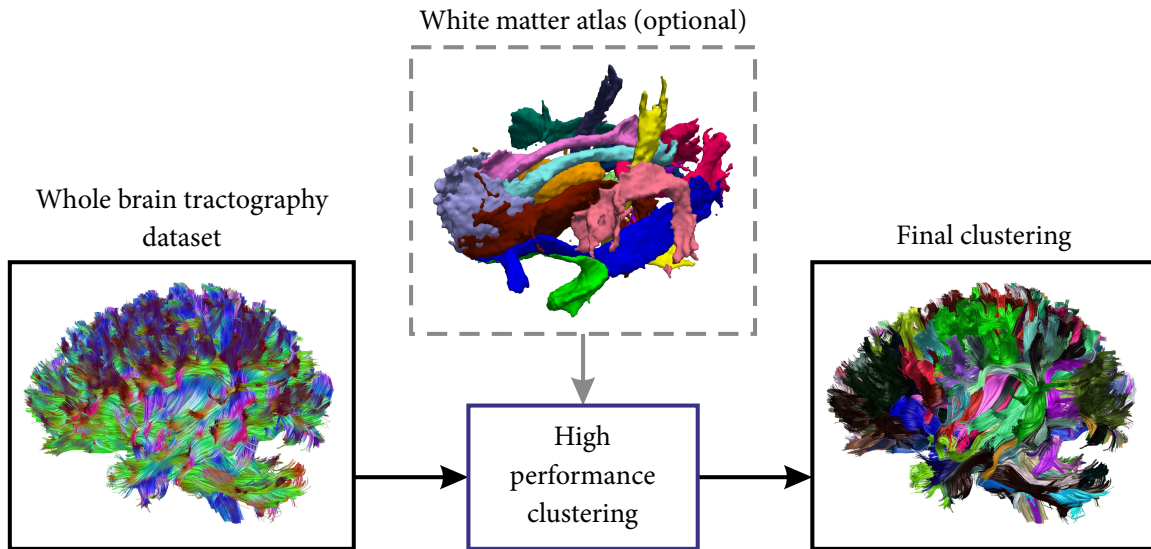


Figure 3.1: Whole brain fiber tractography and fiber tract clustering. Fiber tractography generates streamlines that approximate the underlying axonal pathways of the white matter architecture (left). Tracts are color-coded according to their orientation with red = left-right, green = anterior-posterior and blue = inferior-superior. Clustering methods can then be used to cluster the fiber tracts and to group similar tracts into fiber bundles or sets of tracts (right). Fiber bundles (right) are displayed in distinct colors, and tracts that belong to the same bundle are colored identically. By employing a white matter atlas that consists of several white matter bundles (top middle), automatic extraction can be improved to retrieve anatomically correct fiber bundles.

Hybrid techniques that combine clustering and parcellation-based (or atlas-based) classification will thus be instrumental to move the field of automated fiber tract segmentation forward. For this reason, CATSER is designed to be used in conjunction with a white matter atlas to achieve a more consistent extraction of fiber bundles (see Figure 3.1). With such a predefined segmentation of the white matter, cluster analysis is facilitated in partitioning the tracts according to the predefined regions of the atlas. The additional anatomical information of the atlas is used to guide the clustering by controlling the formation of the clusters on the basis of spatial agreement between fiber tracts and atlas classes.

3.2 Cluster analysis through smartly extracted representatives (CATSER)

3.2.1 Overview

CATSER is based on the CURE (*clustering using representatives*) algorithm that was initially proposed by Guha et al. (2001) for clustering huge databases. Both techniques can essentially be categorized as agglomerative hierarchical clustering methods that use an iterative bottom-up approach in which the most similar tracts are merged during each iteration (see section 2.6).

Tractography datasets can consist of hundreds of thousands of tracts that approximate the axonal pathways in the brain. The number of fiber bundles, however, that are concealed in such datasets is considerably smaller compared to the overall number of tracts. As the number of bundles per dataset is estimated to range between 140 and 500 (Moberts et al., 2005; Visser et al., 2011), it is obvious that each bundle consists of numerous tracts that capture minuscule details of the bundle. Since the tracts in each bundle are mostly similar, the data itself is inherently redundant. CURE and CATSER exploit this redundancy. Instead of clustering the whole dataset as it is the case with conventional clustering methods, they only process a reduced sample to determine a set of prototype bundles. The remaining tracts of the dataset that are not part of the initial sample are then assigned to their most similar prototype clusters.

To reduce computation time and to improve the clustering results, CATSER employs random sampling, partitioning and outlier elimination. Compared to CURE, CATSER performs a modified, more outlier-sensitive clustering to overcome some of CURE's limitations (for details see the article by Karypis et al. (1999)). To this end, the original algorithm was modified to incorporate *local outlier factors* (LOFs) (Breunig et al., 2000) that provide insight into the structural organization of the data. A LOF rates each tract to specify the degree how outlying a tract is with respect to others. The LOFs are used in the cluster analysis to improve the discrimination between true clusters and tracts that are presumed to be outliers (see upcoming section 3.2.5).

3.2.2 Basic CATSER workflow

The processing steps of the CATSER clustering algorithm are presented in this section and are illustrated in Figure 3.2. In order to exploit the redundancy in the data, the whole brain tractography dataset is randomly divided into two sets of tracts (step 1): the reduced random sample and the remaining tracts that are not part of the reduced random sample. The minimum random sample size can be estimated by employing Chernoff bounds (for details see Guha et al. (2001); Xu and Wunsch (2008)), for which it is assumed that every discernible cluster has a minimum size. The minimum reduced random sample size is then computed so that this sample contains at least a tract fraction of each cluster with high probability. As the size of the smallest cluster in the dataset is unknown, the necessary estimation of its size limits the ability to detect smaller clusters if they exist in the dataset.

For various clustering techniques, the computation of similarities between the tracts is integrated into the clustering itself. As this results in redundant computations and degraded performance of the algorithm, the computation of similarities is separated from the cluster analysis. This circumvents not only redundant computations but also enables parallel computation of similarities, which greatly improves the overall performance of the clustering. Similarities as well as LOFs for the sample are therefore precomputed prior to the clustering (step 2).

The subsequent agglomerative hierarchical clustering is used to generate a user-defined number of clusters (details on the clustering process itself are presented in the following section 3.2.3). The clustering is essentially a two-pass process, divided into a partial preclustering and a final clustering stage. At the beginning of the first stage, each tract is considered a singular cluster. The partial preclustering is then primarily a coarse grouping of the most

similar tracts that reduces the number of clusters substantially. A strictly sequential processing of the data during the clustering is therefore not necessary (Guha et al., 2001). Thus, the reduced random sample (from step 1) is randomly divided into a set of partitions each of which contains approximately the same number of fiber tracts (step 3). Each partition is then clustered separately (step 4). In order to speed up the cluster formation during the first pass, the clustering of the separate partitions can be performed in parallel. During this step, unwanted merging of different clusters is avoided by setting the number of clusters that are produced for each partition sufficiently high, and to be at least 2–3 times the user-defined target number of clusters that should be obtained from the entire dataset¹ (see above). After the first pass, the preclustered data of the partitions are joined together without merging clusters (step 5). With this set of joined clusters, the final clustering is performed and prototype clusters are generated (step 6).

During both clustering stages (steps 4 and 6), outlying tracts are identified and removed from the dataset (see section 3.2.4). Due to the shuffling and separation of the data into multiple partitions, tracts of a cluster may be scattered across partitions and tracts may be unintentionally labeled as outliers during the clustering, even if a multitude of similar tracts are adjacent but assigned to other partitions. To warrant that outlying tracts are true outliers, tracts previously labeled as outliers are reevaluated and assigned to the nearest prototype cluster (step 7) if the similarity between tracts and cluster is sufficiently high (for details see section 3.2.7). In the final step, remaining tracts that are not part of the reduced random sample (step 1) and have not been appointed to a cluster yet, are assigned to the nearest prototype cluster (step 8) if the similarity between cluster and tracts is again sufficiently high (see section 3.2.7).

During the whole clustering process, a hierarchical cluster tree that contains all successive merging steps is generated. The cluster tree enables not only visualization of individual clustering steps with dendrograms but also retrospective extraction of bundles or a subset of bundles.

3.2.3 Formation of clusters

Conceptually, CATSER employs agglomerative hierarchical clustering during both clustering stages (steps 4 and 6 in Figure 3.2). Starting with a disjoint set of N clusters $C = \{C_1, \dots, C_N\}$, the iterative clustering process is performed until the number of clusters is reduced to a user-defined target value N_T . In each iteration of the clustering, the two most similar clusters are selected and merged to form a new cluster. For the partial preclustering (step 4), N_T is identical for all partitions.

In order to determine the similarity between two clusters, only a subset of tracts from each cluster is considered. This subset consists of a set of well distributed tracts that capture the geometry of the cluster and act as representative tracts. To start the selection of appropriate representatives tracts, the center of the cluster is determined by locating the cluster medoid.

¹This follows the considerations of Guha et al. (2001).

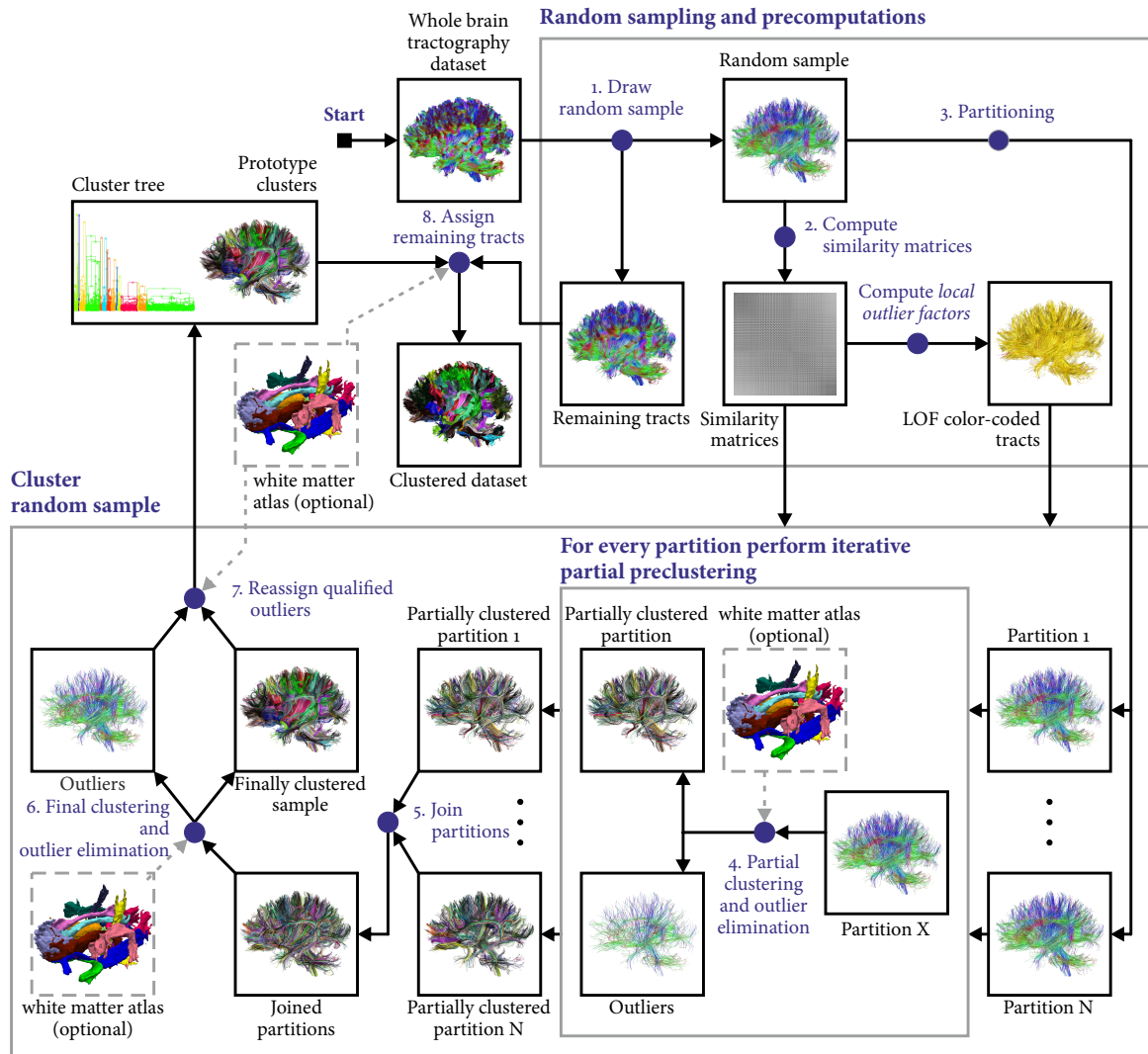


Figure 3.2: Overview of CATSER. The size of the whole brain tractography dataset is reduced by extracting a random sample (1). For this reduced random sample, similarity matrices and *local outlier factors* (LOFs) are computed, which are subsequently used during the clustering process (2). The reduced random sample is automatically divided into a user defined number of N independent partitions (3). A first-pass partial clustering as well as outlier elimination is performed for each partition (4) before resulting clusters of all partitions are joined together (5). Resulting prototype clusters are generated during the second pass of the clustering (6). Outliers are then reassigned to the nearest prototype cluster to obtain the final clustering and a hierarchical cluster tree (7). During the last step, remaining tracts that were not part of the random sample (1) are finally assigned to the nearest prototype cluster (8). The last two steps (steps 7 and 8) are only performed for tracts that are sufficiently similar to a prototype cluster. Other tracts are finally labeled as outliers. To integrate anatomical information into the cluster analysis, a white matter atlas can be used as input during steps 4 and 6–8 (for details see section 3.3).

3.2 Cluster analysis through smartly extracted representatives (CATSER)

For a cluster C with $|C|$ tracts, the medoid m_C is defined as the tract in C whose average distance or dissimilarity to all tracts in C is minimal:

$$m_C = \min_{p \in C} \left(\frac{1}{|C|} \sum_{q \in C} d(p, q) \right). \quad (3.1)$$

After the medoid has been identified, representative tracts are determined iteratively and added to the set of representatives R_C . In each iteration, the tract in C that has the maximum distance to the medoid m_C as well as to all other representatives in R_C becomes a new representative r_{k+1} and is added to $R_C = \{r_1, \dots, r_k\}$:

$$r_{k+1} = \max_{p \in C} \sum_{q \in \{R_C \cup m_C\}} d(p, q) \quad \text{with} \quad R_C = \{r_1, \dots, r_k\} \text{ and } p \notin R_C. \quad (3.2)$$

This iterative selection process guarantees that the representatives are well distributed across the cluster.

The similarity or distances between two clusters C_1 and C_2 is denoted as $d(C_1, C_2)$ and defined to be the distance between the two most similar representatives p, q of C_1 and C_2 . Denoting the sets of representatives for C_1 and C_2 as R_{C_1} and R_{C_2} , the formal definition of $d(C_1, C_2)$ is given by:

$$d(C_1, C_2) = \min_{p \in R_{C_1}, q \in R_{C_2}} d(p, q). \quad (3.3)$$

Representatives play an important role for the formation of clusters by acting as an simplification that approximates the cluster's shape and structure. Since only representatives are used to determine the similarity between clusters, the number of comparisons and the computation time are dramatically reduced. While carefully selected representatives help to prevent negative effects of outlying tracts on the clustering results, such as adverse agglomerative behavior (so called chaining effects, see Xu and Wunsch (2008)), they also enable the technique to cope with clusters of arbitrary shapes. Various conventional clustering techniques (e.g., *k-means*) are limited to spherical cluster shapes in the similarity-based domain. By using representatives to approximate the structural organization of the cluster, the clustering method can handle arbitrarily shaped clusters correctly (Guha et al., 2001).

3.2.4 Outlier handling

In CATSER outliers are handled quite similar to CURE (Guha et al., 2001). By definition (see section 2.6), clusters can be regarded as aggregations of similar tracts. If noise is present, clusters are superimposed by outliers (outlying tracts) which are usually randomly distributed and scattered throughout the dataset. In contrast to real clusters, they are typically isolated and do not exhibit the typical agglomerative behavior (Guha et al., 2001). Compared to tracts that belong to a cluster, the neighborhoods of outlying tracts are generally sparse and distances to other tracts of the dataset are significantly higher.

These observations are translated into the clustering process. An outlier elimination step is performed and clusters that grow very slowly and contain only few tracts at the end of

the clustering are labeled as outliers. If the initial number of clusters is given by N_{ini} and the target number of clusters by N_T , $S_{\text{all}} = N_{\text{ini}} - N_T$ cluster merging steps are performed during the clustering. The outlier elimination is triggered if $S_{\text{all}} \times t$ merging steps remain, where t is user-defined and bounded by $[0, 1]$. During the outlier elimination, the number of tracts per cluster is evaluated for each cluster. Clusters that consist of less than h tracts (user-defined) will be removed from the dataset and their tracts will be marked as outliers.

3.2.5 Local outlier factors (LOFs)

Local outlier factors (Breunig et al., 2000) are a density-based approach to obtain a score for each tract that specifies its outlierness. By employing the precomputed tract-similarities (see section 3.2.2 and step 2 in Figure 3.2), the density distribution of the tracts is analyzed to compute the LOFs. First, the *k-nearest neighbors* (k -NNs, i.e., the k most similar tracts) are determined for each tract. The distances to these k -NNs are then utilized to compute the local tract density for each tract. With the local tract density as well as the local tract densities of its k -NNs, the LOF of each tract is calculated. Practically, a LOF is an estimate on how outlying a tract is compared to its k most similar tracts.

LOFs have the favorable property to specify an outlierness-rating for each tract instead of using a fixed labeling that indicates whether the tract is either an outlier or not. The sparsity of the neighborhood is thereby captured with a single LOF value for each tract. Tracts that lie deep inside of a cluster have a LOF that is approximately 1 or less, whereas the LOF of tracts increases the more isolated the tracts are. Lower² and upper bounds of the LOF depend on the tract-similarities in the dataset. An extensive discussion about the bounds of LOFs and a more detailed description on how to compute LOFs can be found in the original publication by Breunig et al. (2000). Examples of LOFs for a set of points and a fiber bundle are presented in Figure 3.3.

Incorporating LOFs into the cluster analysis of fiber tracts

As LOFs specify a value for each tract that indicates how outlying a tract is compared to its k most similar tracts, it is reasonable to incorporate this structural information into the clustering to support the outlier handling. The outlierness information provided by the LOFs is used as complementary information besides the similarities to adjust the pairwise similarities between tracts during the formation of clusters. Here, two scenarios have to be considered.

On the one hand, intra-cluster distances have to be considered. They are primarily related to the determination of the cluster representatives, which are iteratively selected by finding those tracts in the cluster that maximize the distance to all other tracts in the cluster. On the other hand, inter-cluster distances have to be considered as well as they are used to determine those clusters that will be merged in each iteration of the clustering. In the latter case, the two clusters that minimize the distance between them have to be found (see Eq. 3.3). These two scenarios are obviously conflicting. While the distance between tracts should be maximized in

²For LOFs, the (theoretical) greatest lower bound is zero (Breunig et al., 2000).

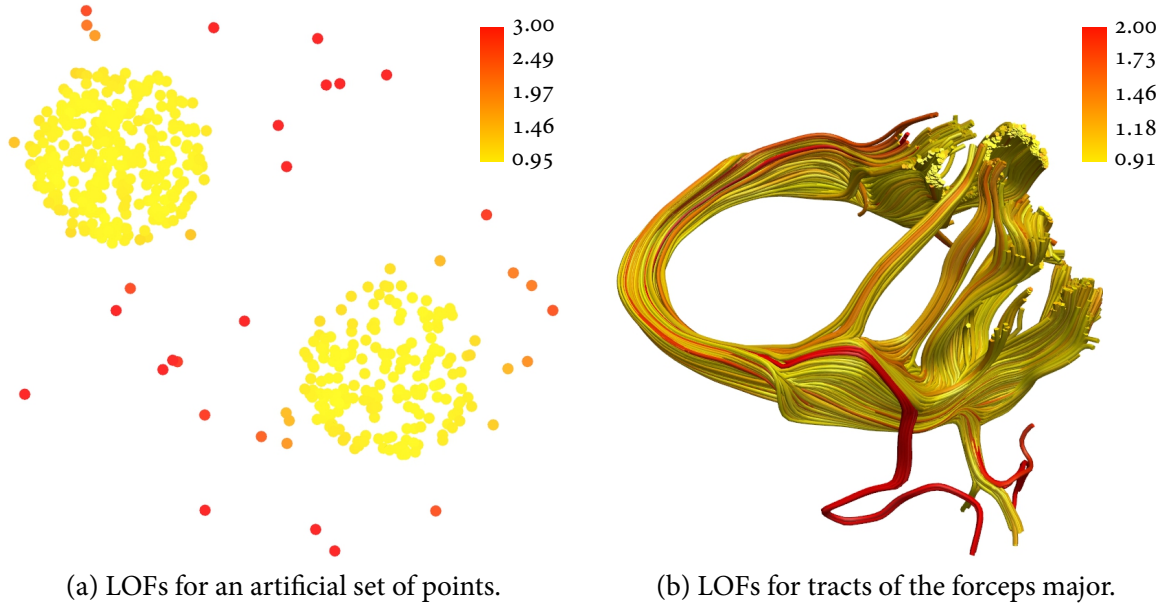


Figure 3.3: Color-coded visualization of *local outlier factors* (LOFs) for a set of points in 2D (a) and for tracts of a fiber bundle (b). LOFs were computed to indicate how outlying the points and tracts are compared to their k most similar neighbors. For both examples, k was set to 15. Points and tracts are color-coded with the values of their respective LOFs, where yellow denotes low and red high LOF values. To highlight the variability of the LOFs, the maximum range of coloring is limited to a LOF of 3 in (a) and a LOF of 2 in (b). The maximum LOF is 5.34 in (a) and 4.48 in (b).

the first scenario, it should be minimized in the second scenario. Consequently, both scenarios have to be handled differently (see below).

However, irrespective of the actual scenario, a LOF-based correction factor $\text{corr}(p, q)$ is initially defined to adjust the pairwise similarities between two tracts p and q . By denoting the *local outlier factor* of p and q with $\text{LOF}(p)$ and $\text{LOF}(q)$ respectively, $\text{corr}(p, q)$ is defined as follows:

$$\text{corr}(p, q) = \left(\frac{\text{LOF}(p) + \text{LOF}(q)}{2} \right)^2. \quad (3.4)$$

As long as p, q are not outlying and their LOFs are around 1, $\text{corr}(p, q)$ yields ≈ 1 . If either p, q or both are outlying, the LOFs are elevated and the correction factor for p and q will increase by $((\text{LOF}(p) + \text{LOF}(q))/2)^2$.

Adjusting intra-cluster distances

Intra-cluster distances are used to determine the cluster medoid and to select tracts of the cluster that act as representatives. However, the representative selection algorithm promotes

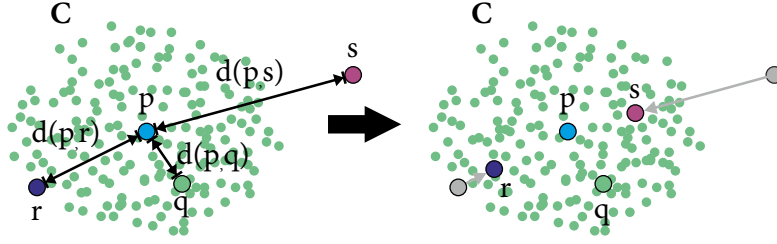


Figure 3.4: Influence of LOFs on intra-cluster distances. Given one cluster C and the set of points $\{p, q, r, s\} \in C$, the influence of LOFs on the intra-cluster distances between p and the exemplary points q, r, s is unveiled. In the example, the LOF of p, q is approximately 1 whereas the LOF of r is slightly increased and the LOF of s is considerably elevated. As large LOFs result in reduced distances between points (attraction effect), the LOF-corrected distance between p, s is considerably reduced, while the correction only slightly reduces the distance between p, r . Since p, q are not outlying (LOF ≈ 1), the LOF correction has almost no effect on the distance between p, q .

the selection of outlying tracts as representatives³. As this might cause inferior clustering results (due to chaining), careful selection of representatives is essential. Intra-cluster distances are therefore adjusted to prevent outlying tracts from being selected. To assess the similarity between tracts p and q , the LOF-corrected distance $d'(p, q)$ is thus used instead of $d(p, q)$:

$$d'(p, q) = d(p, q) \times \frac{1}{\text{corr}(p, q)}. \quad (3.5)$$

The inverse application of the correction factor $\text{corr}(p, q)$ (Eq. 3.4) essentially applies a spatially dependent attraction that pulls tracts with an increased LOF into the cluster. If tracts have a large LOF, the correction factor $\text{corr}(p, q)$ will also be larger and the distance $d(p, q)$ will decrease. On the contrary, if tracts have a LOF that is ≈ 1 , tracts are practically not affected. Due to this effect, tracts with a large LOF will suffer a penalty and have a reduced distance to the medoid and to other representatives (see Figure 3.4 for an example with an artificial set of points in 2D). This decreases the possibility that outlying tracts with a large LOF are selected as representatives of a cluster.

Adjusting inter-cluster distances

To adjust the distance between two different clusters C_1 and C_2 (see Eq. 3.3), the distance $d(p, q)$ between their two most similar representative tracts p, q is weighted with the LOF-based correction $\text{corr}(p, q)$. Eq. 3.3 is thus replaced with the new LOF-adjusted distance $d_{\text{corr}}(C_1, C_2)$:

$$d_{\text{corr}}(C_1, C_2) = \min_{p \in R_{C_1}, q \in R_{C_2}} d''(p, q), \quad (3.6)$$

$$d''(p, q) = d(p, q) \times \text{corr}(p, q). \quad (3.7)$$

³During the representative selection, tracts are selected that have the largest distance to the cluster medoid and previously selected representatives of the cluster (see section 3.2.3).

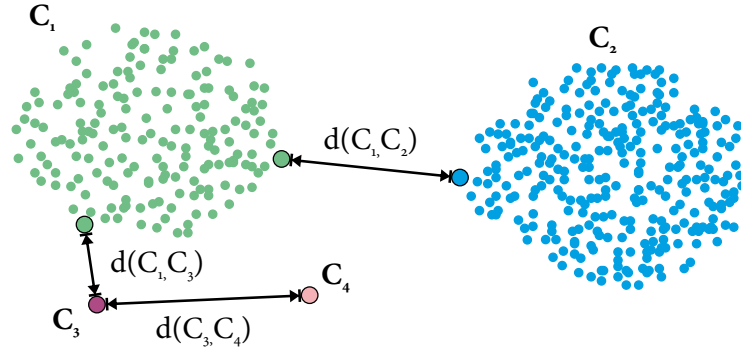


Figure 3.5: Influence of LOFs on inter-cluster distances. Given clusters C_1 – C_4 and the shortest distances $d(C_1, C_2)$, $d(C_1, C_3)$, $d(C_3, C_4)$ between the clusters, LOF-related effects on the inter-cluster distances are emphasized. Depending on the cluster's surroundings, the LOFs for points in C_1 and C_2 are ≈ 1 , whereas the LOF for points in the solitary clusters C_3 , C_4 is increased and larger than one. Hence, the LOF-based correction has almost no effect on the distance between clusters C_1 , C_2 . As a reciprocal relation is used for the computation of inter-cluster distances compared to intra-cluster distances, the distance between C_1 and C_3 will increase due to C_3 's high LOF. This repulsion effect is even more pronounced between clusters C_3 and C_4 , because points in both clusters have high LOF values.

This correction affects only the distances between the clusters and influences when and which clusters are merged. In order to understand the mechanisms behind the correction, the clustering should be considered as a continuous process in which clusters are not yet finished but successively formed. If the LOF for the representatives of two distinct clusters is ≈ 1 , the distance between the clusters is not altered. By definition, these representatives need to belong to true clusters and are located inside the clusters (otherwise the LOF would not be ≈ 1). If representatives are located at the boundaries of a cluster (cmp. Figure 3.3), their LOF is slightly increased. As a result, the clusters will experience minimum repulsion and will be clustered slightly later due to their increased distances. If either one or both representatives of the clusters are outliers, they will possess a high LOF. As a result of the LOF-correction, the distance between the clusters will considerably increase and they will be merged substantially later (if merged at all). An exemplary illustration that depicts the adjustment of the inter-cluster distances for an artificial set of points in 2D is given in Figure 3.5.

This correction effectively contributes to the employed outlier elimination strategy, which is based on the assumption that outlying tracts will be far more isolated than other tracts. As a result of the LOF correction, outlying tracts or clusters will become even more isolated and will therefore not be clustered at all or only at the very end.

3.2.6 Choosing the number of representatives

An important aspect for the selection of representatives (see section 3.2.3) is the number of representatives that are used for clustering. In practice, this number is a trade-off between correct clustering, accurately assessing the cluster shape, achieving computational efficiency and robustness to noise (see Figure 3.6). If only one representative is used, robustness against

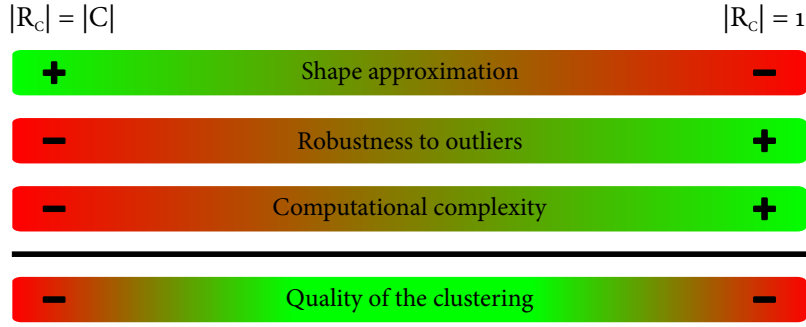


Figure 3.6: Influence of the number of representatives $|R_C|$ on the clustering process for a cluster C . If only one representative is used ($|R_C| = 1$, right side), the shape and extent of cluster C cannot be assessed accurately, whereas robustness to noise and computational complexity are positively affected. In contrast, if all elements of a cluster are selected as representatives ($|R_C| = |C|$ with C being the number of tracts in cluster C , left side), shape and extent of the cluster are assessed accurately, while susceptibility to noise and computational complexity both increase. In practice, the selection of the number of representatives is a trade-off between correct clustering, accurately assessing cluster shape, achieving computational efficiency and robustness to noise. In the illustration, positive effects are shown in green, whilst negative effects are depicted in red.

noise as well as computational efficiency is achieved but the shape of the cluster cannot be described accurately, which will result in degraded clustering quality. On the contrary, if all tracts of the cluster are employed as representatives, the cluster shape will be preserved accurately but computation time will increase. In addition, potential outliers will also be included in the list of representatives. This will impair CATSER's ability to deal with outliers since CATSER is based on the notion to prevent the selection of outliers as representatives during the clustering. In order to obtain optimal clustering results, it is therefore necessary to find a balance between both extreme cases.

Even though Figure 3.6 implies a static correspondence between the number of representatives and cluster size, the clustering itself is a dynamic process, through which clusters evolve and grow. During the formation of a cluster the number of representatives $|R_C|$ is therefore chosen with respect to the individual size of that particular cluster C . In order to select $|R_C|$, the cluster formation process is divided into two stages (see Figure 3.7). In the first stage, $|R_C|$ is chosen with respect to the individual size of the cluster and adapted to reflect changes in cluster size. As the cluster size and the number of representatives increase the computational efficiency decreases due to the additional calculations that have to be performed for each additional representative. In order to maintain the computational efficiency, a second stage with a constant number of representatives is defined. To select the number of representatives, a monotonically increasing, piecewise-defined function is used that describes these two stages (see Figure 3.7). The function $f(x)$ computes the number of representative $|R_C|$ depending on of the number of tracts in the cluster $x = |C|$ and the user-defined values s and x_1 :

$$|R_C| = f(x) = \begin{cases} s \times x & \text{if } 1 \leq x \leq x_1, \\ s \times x_1 & \text{if } x > x_1 \end{cases} \quad (3.8)$$

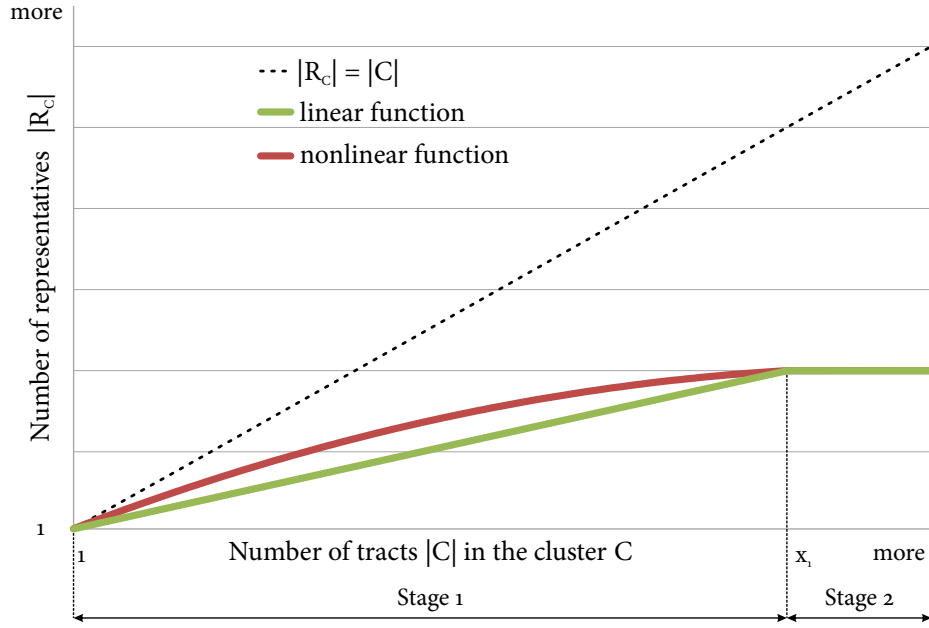


Figure 3.7: Determination of the number of cluster representatives. To determine the number of representatives $|R_C|$ for a cluster C a two stage approach is used. As long as the number of tracts $|C|$ in cluster C is smaller than x_1 , R_C is selected depending on $|C|$ using either a linear or a nonlinear function (stage 1). If $|C|$ is larger than x_1 , R_C is set to a constant value (stage 2).

with $s \in]0, 1]$ and $|R_C| \geq 1$ for all x . Instead of such a linear function, nonlinear functions (e.g., monotonically increasing, interpolation functions or square root functions) might also be employed to achieve a smooth transition between the stages.

3.2.7 Labeling and reassigning tracts

In CURE and CATSER, slowly growing clusters that contain only few tracts at the end of the clustering are labeled as outliers. Due to the randomized splitting of the data into multiple partitions during the clustering, tracts may be unintentionally labeled as outliers, even if a multitude of tracts are spatially located nearby in other partitions. In order to correct for possibly wrong assignments, outliers are reevaluated. The distances to all extracted clusters are assessed and outliers are reassigned to the most similar cluster if the similarity between outlier and cluster is high enough.

By treating the outlying tract o as a singular cluster C_o that consists only of tract o , the similarity $d(C_o, C_i)$ to all N regular clusters $C = \{C_1, \dots, C_N\}$ is computed. Let C_M be the most similar cluster from C to C_o :

$$C_M = \min_{C_i \in C} d(C_o, C_i), \quad (3.9)$$

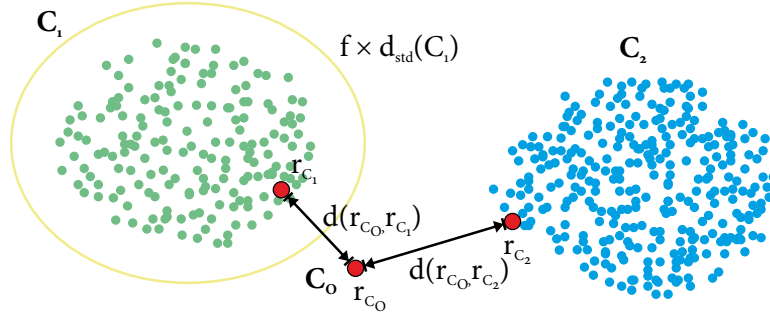


Figure 3.8: Assignment and reassignment of tracts in CATSER. Given the prototype clusters C_1 , C_2 , the solitary cluster C_o and a subset of representatives, the closest cluster to C_o is C_1 since $d(r_{C_o}, r_{C_1}) < d(r_{C_o}, r_{C_2})$. The representatives with the shortest distance are denoted as r_{C_o} and r_{C_1} . C_o is then (re-) assigned to the closest cluster C_1 only if the distance $d(r_{C_o}, r_{C_1})$ is smaller than $f \times d_{\text{std}}(C_1)$, where $d_{\text{std}}(C_1)$ is the standard deviation of the distances between all representatives in C_1 .

and r_{C_o} , r_{C_M} the two closest representatives of C_o and C_M respectively. If $d_{\text{std}}(C_M)$ is the standard deviation of the distances between all representatives in C_M and the inequality

$$d(C_o, C_M) \leq f \times d_{\text{std}}(C_M) \quad (3.10)$$

is satisfied, tract o is assigned to C_M (see Figure 3.8). Otherwise, the status of o as outlier is maintained. The factor f can be chosen arbitrarily by the operator (with $f > 0$) and is used to regulate how similar tracts need to be in order to permit the reassignment.

Subsequent to the clustering and the reassignment of wrongly labeled tracts, the set of tracts C_{out} that were not part of the initial random sample C_{in} (see section 3.2 and step 1 in Figure 3.2) have to be processed. Hereby, they are either assigned to the nearest cluster or labeled as outlier. This processing step is carried out in a similar way as the reassignment described above. For each tract $o \in C_{\text{out}}$ the distance to all regular clusters is computed and tract o is assigned to the nearest cluster if inequality 3.10 is satisfied. As the computation of the LOFs for the entire dataset is too time consuming⁴, it is assumed that the LOF is 1 for each tract in C_{out} . The labeling of the tracts can be realized independently, which enables the separation of the task into mutually independent subtasks. Parallel processing can therefore be employed to simultaneously process the tracts in C_{out} and to reduce the computation time.

In order to choose different values for f (see inequality 3.10), the factor f_r is used for the reassignment of outlying tracts and f_l for the labeling of remaining tracts.

3.2.8 Parameter selection

The selection of adequate parameters for the clustering is essential to achieve good clustering results. Important clustering parameters are therefore reviewed and discussed along with general guidelines for choosing parameters. An exhaustive list of parameters with a description,

⁴Note that the complete similarity matrix has to be available to compute the LOFs for the entire dataset.

3.2 Cluster analysis through smartly extracted representatives (CATSER)

recommended and default values for each parameter is presented in Table 3.1. Nevertheless, as every dataset is unique and contains different tracts, adjusting certain parameters might be necessary.

Cluster generation

For the cluster generation with CATSER various parameters have to be set up (see sections 3.2.2–3.2.7). Partitions are primarily used to reduce the computation time by splitting a large sample into smaller parts that are separately clustered. However, if the number of partitions is too high ($N_p > 50$), it is likely that the partitions contain too few tracts to perform a correct clustering (Guha et al., 2001). To prevent a negative effect on the clustering, the number of partitions should be kept sufficiently small (1–10).

Setting the number of representatives is straightforward. The number of representatives of a cluster has to be lower than the number of tracts in the cluster to guarantee that the representative selection process works (see section 3.2.6). As the number of representatives affects the computational load, a low number of representatives (e.g., no more than 40 representatives for clusters with more than 120 tracts) preserves the computational efficiency and reduces the processing time. Nevertheless, if the number of representatives is too low to adequately preserve the shape of the cluster, clustering may be affected.

During the clustering procedure, the outlier elimination parameters t and h influence when the outlier elimination is performed and up to which size a cluster is regarded an outlier (see section 3.2.3). If the outlier elimination is performed too early during the clustering, tracts that belong to clusters will be wrongly removed. If it is performed too late, outliers will be unintentionally integrated into the clusters.

The outlier elimination during the first stage should be performed at the very end of the preclustering, where clusters should be labeled as outliers if they contain only 1–2 tracts. The outlier elimination step of the second stage should be performed earlier and for clusters that consist of up to 3–7 tracts. Ultimately, however, the optimal choice of outlier elimination parameters is governed by the amount of noise that is in the dataset. If a large amount of noise is present, a more stringent outlier elimination is inevitable to achieve adequate clustering results.

Reassignment and labeling

The multipliers f_r and f_s are used to regulate how similar tracts have to be in order to be assigned to a cluster (see section 3.2.7). The smaller the values the more similar the tracts have to be in order to be assigned to a cluster. The values for f_r and f_s can be chosen arbitrarily, but values between 1 and 2 usually work quite well.

However, care should be exercised with the setting of f_r , since it governs the reassignment of tracts that were already marked as outlier. If f_r is set to high ($f_r > 1$), true outliers might be integrated into the clusters and the separation between clusters decreases.

Table 3.1: Overview of clustering parameters.

Name	Description/Role of the parameter	Reasonable values	Default values
N_{in}	size of the sample	varies	–
k	number of <i>k-nearest neighbors</i> (LOFs)	$10 - C_N^5$	15
N_T	target number of clusters	varies	250
N_p	number of partitions	1–30	3
χ_1	adjust the number of representative	varies (40–200)	120
s	adjusts the number of representatives (slope of stage 1)	0.1–0.9	0.3
t_1	outlier elimination trigger (1 st clustering stage)	0.05–0.2	0.1
h_1	outlier elimination threshold (1 st clustering stage)	1–2	2
t_2	outlier elimination trigger (2 nd clustering stage)	0.1–0.3	0.2
h_2	outlier elimination threshold (2 nd clustering stage)	3–7	3
f_r	multiplier for the reassignment of outlying tracts	1–2	1
f_l	multiplier for the labeling of unprocessed tracts	1–2	1.5

LOF computation

The LOF of a tract is solely based on the number of k -NNs that will be used to assess the local tract density. Since the value of k has a direct impact on the LOFs, Breunig et al. (2000) derived guidelines for reasonable values of k . To achieve a stable solution without statistical fluctuations, it is sufficient to choose $k \geq 10$. Beyond that, k implicitly defines the minimum cluster size. Consider the smallest cluster C in a dataset with $|C|$ tracts. If $k > |C|$, the k -NNs of each tract in cluster C include not only tracts from the cluster C itself but also tracts that belong to another cluster or tracts that are outliers.

In this case, LOFs will be artificially elevated. In order to prevent an unintentional increase of the LOFs for small clusters, the value of k should not be larger than the smallest cluster expected in the dataset: $10 \leq k \leq |C|$.

3.2.9 Differences to CURE

The original CURE algorithm was initially proposed by Guha et al. (2001) for the clustering of huge databases. It employs random sampling, partitioning and outlier elimination to enhance the cluster analysis. During the clustering, CURE computes a centroid and uses a fixed number of representatives for all clusters. During the formation of clusters, representatives are

⁵ C_N is the expected minimum cluster size.

shrunk towards the cluster centroid with a fixed value to reduce chaining and to dampen effects of outliers.

Both a fixed number of representatives as well as a fixed shrink factor for all clusters have drawbacks (Karypis et al., 1999). As clusters vary in shape and extent, more complex cluster geometries cannot be described properly by using the same number of representatives that are sufficient for simple cluster shapes. Shrinking of clusters preserves the shape but fundamentally changes the extent of the clusters. Depending on other clusters and outliers that are located in the vicinity of a cluster and its representatives, the optimal magnitude of the shrinking may differ for different clusters and their representatives. The shrinking is thereby usually not identical but different for distinct clusters and representatives. Representatives with none or merely a few outlying tracts in their vicinity only need minimum shrinking, whereas representatives with many outliers in the neighborhood require a higher shrink factor.

CURE has been applied to a wide range of applications, but it cannot be used for the cluster analysis of non-Euclidean similarity measures. This is due to its utilization of concepts that are solely available in Euclidean metric spaces (e.g., centroids, averaging, shrinking). In contrast to CURE, CATSER is specifically designed to handle non-Euclidean similarity measures. The computation of similarities and the cluster analysis is thereby strictly separated, which enables precomputation of similarities and LOFs.

For the clustering, CATSER uses a LOF-based strategy to select more qualified representatives that are not outlying. This strategy is conceptually related to CURE's shrinking approach but uses spatially dependent factors to adjust the distances between tracts of a cluster, hereby influencing the selection of representatives (see section 3.2.5).

As CATSER is based on the notion to prevent selection of outlying tracts as representatives, a fixed number of representatives will have a negative effect on outlier elimination and can therefore not be used⁶. If the number of representatives is fixed and set as high as the number of tracts, all tracts will become representatives even if a tract is outlying. In order to circumvent this behavior, CATSER adjusts the number of representatives individually for each cluster with respect to the number of tracts in the cluster (see section 3.2.6). This allows the selection algorithm to reject outlying tracts.

3.3 Integration of anatomical information into the cluster analysis

In order to incorporate anatomical information into the cluster analysis, a white matter atlas that contains various fiber bundles can be instrumented. As the clustering is based on the

⁶If the number of representatives is fixed and the number of tracts in a cluster is smaller than the number of representatives, each tract in the cluster will become a representative regardless whether the tract is an adequate representative or not.

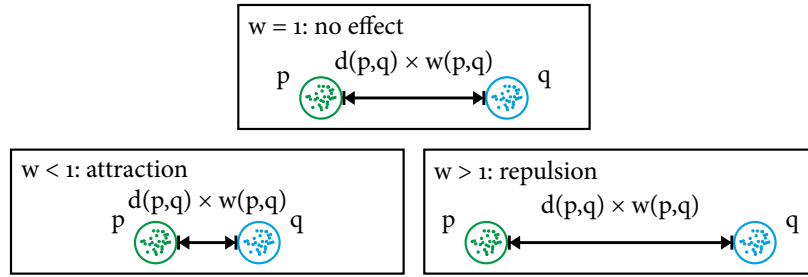


Figure 3.9: Effects of different weighting factors. The weighting factors guide the clustering by modulating the distance between the clusters p and q according to their anatomical correspondence in the atlas. While a weighting factor $w(p, q) = 1$ has no effect, a weighting factor $w(p, q) < 1$ will decrease the cluster distance (attraction). Contrary, a weighting factor $w(p, q) > 1$ will increase the distance between the clusters (repulsion).

principle of merging clusters with the highest similarity, an effective way to influence the clustering is to modulate the distance $d(C_1, C_2)$ between the clusters C_1 and C_2 :

$$d(C_1, C_2) = d(C_1, C_2) \times w(C_1, C_2). \quad (3.11)$$

As those clusters with the shortest distance are merged in each step, the weighting factor $w(C_1, C_2)$ has a direct impact on the clustering and how clusters are formed. While values of $w(C_1, C_2) > 1$ will increase the distances and lead to a repulsion, a weighting $w(C_1, C_2) < 1$ will result in attraction for C_1, C_2 (see Figure 3.9). This effect is exploited and the weighting $w(p, q)$ is determined with respect to the atlas-class membership of the clusters p and q . To guide the cluster analysis, this weighting is assessed and applied in each successive step of the clustering as well as in the reassignment and the labeling stage. The steps in which the atlas has to be employed for the atlas integration are depicted in Figure 3.2 (steps 4 and 6–8), showing the atlas as optional input in grey dotted boxes.

3.3.1 The white matter atlas

For the cluster analysis a voxel-based white matter atlas, constructed with a semi-automatic method was applied (see section 6.2 that describes the atlas generation in detail). The atlas consists of various white matter (WM) structures (i.e., the classes of the atlas) and each WM structure contains a set of voxels that belong to the structure. Hereby, different regions can overlap and voxels are allowed to belong to multiple classes of the atlas. In Figure 3.1, a white matter atlas is shown with a selection of fiber bundles in distinct colors.

3.3.2 Determining the class membership for tracts

In order to determine the weighting factor $w(p, q)$, the atlas-class membership for the individual tracts and the clusters (groups of tracts) has to be defined. The class-membership of single tracts is computed by rasterizing each tract and determining the spatial agreement of the tract and the classes in the atlas (Figure 3.10).

3.3 Integration of anatomical information into the cluster analysis

For each atlas class A and each tract t , the spatial agreement $s(t, A)$ is the ratio between the number of voxels N_I that intersect the class in the atlas and the number of all tract-voxels N

$$s(t, A) = \frac{N_I}{N}. \quad (3.12)$$

For a cluster C with a set of $|C|$ tracts, the atlas-class membership to class A is computed by averaging the spatial agreement between all tracts in the cluster and class A :

$$s(C, A) = \frac{1}{|C|} \sum_{t \in C} s(t, A). \quad (3.13)$$

3.3.3 Determining the weighting factor for two clusters

During the clustering, the weighting has to be determined and applied in each successive step. For two clusters C_1 and C_2 , the two classes of the atlas A_{C_1} , A_{C_2} with the highest spatial

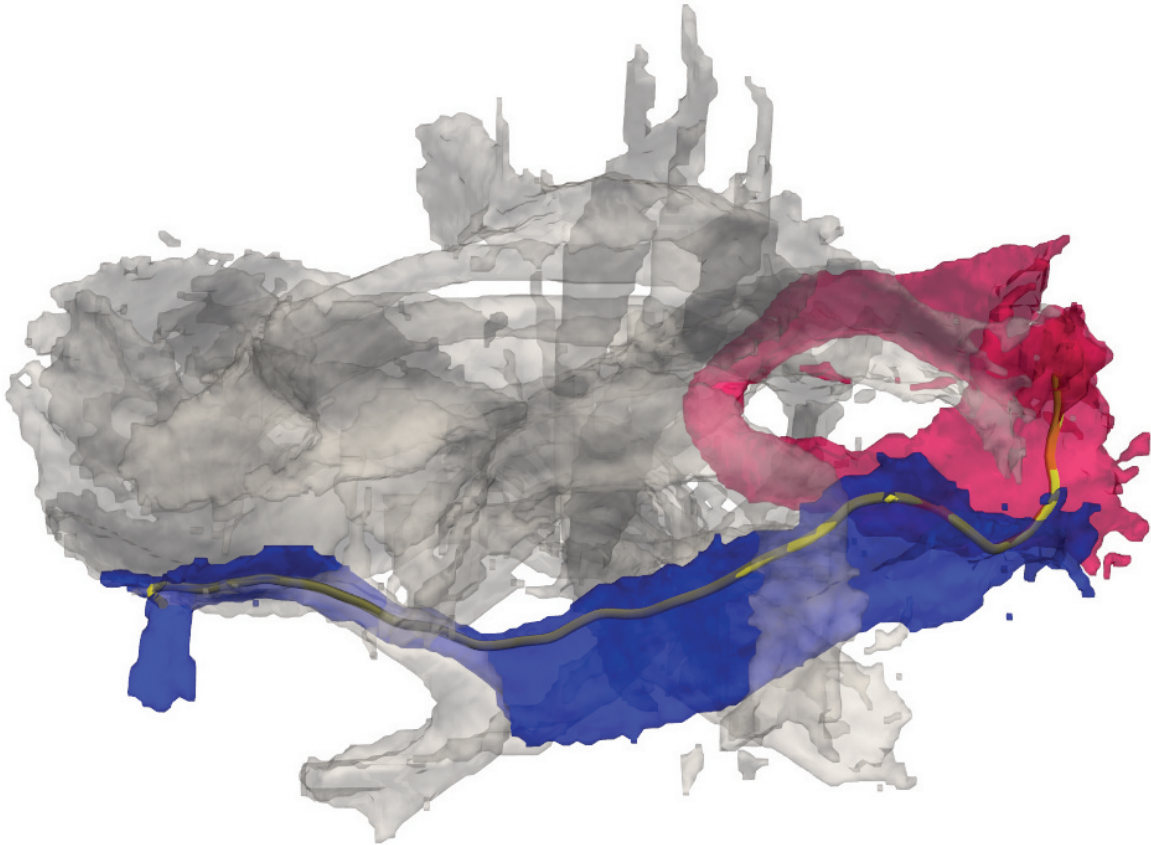


Figure 3.10: The figure shows a single fiber tract that traverses through two classes of an atlas. A single fiber tract (yellow) traverses through the inferior fronto-occipital fasciculus (IFO, in blue) and the forceps major (Fmaj, in red). The remaining atlas classes that share no spatial volume with the tract are displayed in grey. The class membership of the tract for the IFO is ≈ 0.73 and for the Fmaj ≈ 0.12 .

agreement $s(C_1, A_{C_1})$, $s(C_2, A_{C_2})$ are identified. For improved readability, the short notation s_{C_1} and s_{C_2} will be used as abbreviation. Then, one can distinguish between four cases:

- case 1: cluster C_1 and C_2 have no corresponding class in the atlas,
- case 2: either cluster C_1 or cluster C_2 has a corresponding class in the atlas (s will be the spatial agreement for the non-empty class),
- case 3: both cluster C_1 , C_2 correspond to the same class in the atlas,
- case 4: both cluster C_1 , C_2 correspond to different classes in the atlas.

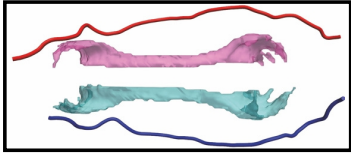
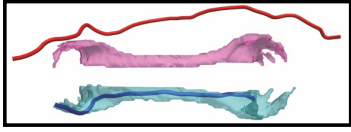
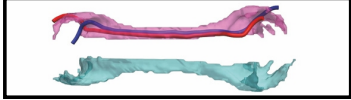
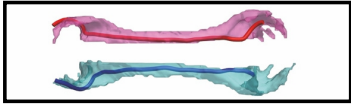
Each case modulates the weighting in a different way:

$$w(p, q) = \begin{cases} 1, & A_{C_1} = \emptyset \wedge A_{C_2} = \emptyset \text{ (case 1),} \\ 1/(1 - \frac{s}{2}), & A_{C_1} = \emptyset \oplus A_{C_2} = \emptyset \text{ (case 2),} \\ \sqrt{\frac{(1-s_{C_1})^2 + (1-s_{C_2})^2}{2}}, & A_{C_1} = A_{C_2} \text{ (case 3),} \\ 1/\sqrt{\frac{(1-s_{C_1})^2 + (1-s_{C_2})^2}{2}}, & A_{C_1} \neq A_{C_2} \text{ (case 4).} \end{cases} \quad (3.14)$$

The four cases are summarized in Table 3.2. For the sake of simplicity, the clusters in the table are shown as singular clusters with only one single tract. However, as clusters are merged during the iterative clustering procedure, the cluster will grow in size and can thus be composed of multiple tracts. In this case, the spatial agreement between the cluster and the atlas classes is the average of the spatial agreement between all tracts and the atlas class (see Eqs. 3.12 and 3.13).

3.3 Integration of anatomical information into the cluster analysis

Table 3.2: The four cases that determine the weighting factor for the atlas guidance are shown. In order to demonstrate and visualize the four cases, the classes of the atlas that correspond to the left and right cingulum bundle are shown in pink and cyan along with two tracts that represent two clusters (shown in red and blue).

case	description	illustration	weighting	effect
1	both tracts have no correspondence in the atlas ($A_{C_1} = \emptyset \wedge A_{C_2} = \emptyset$)		$w(p,q) = 1$	none
2	one tract has a correspondence in the atlas ($A_{C_1} = \emptyset \oplus A_{C_2} = \emptyset$)		$w(p,q) > 1$	repulsion
3	both tracts correspond to the same class in the atlas ($A_{C_1} = A_{C_2}$)		$w(p,q) < 1$	attraction
4	both tracts correspond to different classes in the atlas ($A_{C_1} \neq A_{C_2}$)		$w(p,q) \gg 1$	repulsion

4

Chapter 4

Similarity measures for fiber tracts

I paint objects as I think them, not as I see them.

— Pablo Picasso (1881–1973)

Features and similarity measures are an integral part of cluster analysis and define the criteria that are used to achieve a distinguishable separation between tracts. While a variety of different similarity measures are integrated into the presented clustering framework, it is not purposeful to discuss all of them. Instead, only those features and similarity measures that were used for the experiments and analyses in this thesis are formally defined and reviewed.

For a selection of similarity measures that were used by other authors for clustering fiber tracts, the interested reader is referred to the publications of Ding et al. (2003); Moberths et al. (2005); Batchelor et al. (2006); Maddah et al. (2008); Visser et al. (2011) and Garyfallidis et al. (2012).

4.1 Overview

Cluster analysis of fiber tracts is not only based on the clustering algorithm itself, but also depends on the features and similarity measures that are employed to estimate the similarity between tracts. Challenges of cluster analysis of fiber tracts include not only the amount of tracts that have to be compared during the analysis, but also the high intrinsic complexity of fibers tracts. As tracts are an ordered sequence of points in 3D space, the trajectories of the tracts are defined by the sequential arrangement of tract points (cmp. section 2.4). In principle, the tract points represent the most fundamental feature, and the generation of new, unique features is therefore predominately (but not exclusively) based on them (cmp. section 2.6, feature extraction).

Due to the inherent complexity of fiber tracts, it is not trivial to determine the correspondence between tracts. Though various simplistic features can be extracted and employed to determine tract similarity, they are mostly unable to handle the intrinsic complexity that is

concealed in the formation of successive points¹. As advanced similarity measures are usually better qualified to handle the profound complexity of fiber tracts² and to capture their complex structural organization, this chapter primarily focuses on more advanced similarity measures.

In the following sections, various similarity measures that were used for clustering experiments in this thesis are defined. These include mostly complex similarity measures, such as the *minimum tract distance* (MD), the *Hausdorff distance* (HD), the *matched point distance* (MPD) and the *combined distance* (CD), but also more simple similarity measures (*tract centroid distance* (TCD) and *tract orientation similarity* (TOS)).

4.2 Tract centroid distance measure (TCD)

To determine the similarity of two tracts p, q with respect to the distance between their tract centroids, the *tract centroid distance* $d_{\text{TCD}}(p, q)$ is defined as the Euclidean distance between the two centroids \mathbf{G}_p and \mathbf{G}_q :

$$d_{\text{TCD}}(p, q) = \|\mathbf{G}_p - \mathbf{G}_q\|. \quad (4.1)$$

As the tract points are not necessarily distributed equally, a simple averaging of tract point coordinates would inadvertently shift the centroid. In order to cope with unequally distributed points along an arbitrary tract p described by a finite set of N points, p 's center of gravity \mathbf{G}_p is used to compute the centroid similarity. If two successive points \mathbf{p}_i and \mathbf{p}_{i+1} define an individual segment of p , the centroid and the length for each segment of p is denoted by \mathbf{g}_i and l_i respectively (see Figure 4.1):

$$\mathbf{g}_i = \mathbf{p}_i + \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{2}, \quad (4.2)$$

$$l_i = \|\mathbf{p}_{i+1} - \mathbf{p}_i\|. \quad (4.3)$$

The center of gravity for tract p can now be defined as the length-weighted average of all segment centroids \mathbf{g}_i of p . If L_p is p 's overall tract length:

$$L_p = \sum_{i=1}^{N-1} l_i, \quad (4.4)$$

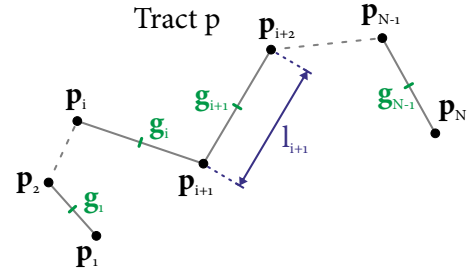


Figure 4.1: Illustration of a fiber tract p with N tract points (in black). Two successive points \mathbf{p}_i and \mathbf{p}_{i+1} define an individual segment (in grey). Each segment s_i has a segment centroid \mathbf{g}_i (in green) and a length l_i (in blue, only one segment length is shown).

¹Consider, for example, a similarity measure that assesses only the similarity between tract centroids and neglects tract orientation. If fiber tracts have similar centroids but are perpendicular to each other, they will still be partitioned into the same cluster due to the fact that the orientation as an important tract characteristic has not been considered.

²Nevertheless, it should be noted that the increasing complexity of features and similarity measures also increases the computational complexity as more operations are involved to determine tract similarity.

\mathbf{G}_p is given by

$$\mathbf{G}_p = \frac{1}{L_p} \sum_{i=1}^{N-1} \mathbf{g}_i \times \mathbf{l}_i. \quad (4.5)$$

4.3 Tract orientation similarity measure (TOS)

To assess the similarity between two tracts p , q according to their orientation, the *tract orientation similarity* $d_{\text{TOS}}(p, q)$ is defined as the angular distance between their orientation vectors \mathbf{O}_p and \mathbf{O}_q :

$$d_{\text{TOS}}(p, q) = \arccos \left(\frac{\mathbf{O}_p \bullet \mathbf{O}_q}{\|\mathbf{O}_p\| \times \|\mathbf{O}_q\|} \right). \quad (4.6)$$

As main orientation vector \mathbf{O}_p and \mathbf{O}_q the endpoint-orientation of the two tracts is used (Wang et al., 2007):

$$\mathbf{O}_p = \mathbf{p}_N - \mathbf{p}_1, \quad (4.7)$$

$$\mathbf{O}_q = \mathbf{q}_M - \mathbf{q}_1. \quad (4.8)$$

4.4 Minimum tract distance (MD)

The *minimum tract distance* $d_{\text{MD}}(p, q)$ between two arbitrary tracts p , q with a finite set of N , M points respectively, is defined as the distance between the two closest points \mathbf{p}_i , \mathbf{q}_j of p and q :

$$d_{\text{MD}}(p, q) = \min_{\mathbf{p}_i \in p, \mathbf{q}_j \in q} \|\mathbf{p}_i - \mathbf{q}_j\|. \quad (4.9)$$

Even though the MD is computational demanding, it is a simplistic similarity measure that is of limited use as it neglects the orientation and the course of the tracts.

4.5 Hausdorff distance measure (HD)

A popular similarity measure that determines the distance between two finite sets of points is the *Hausdorff distance* (Rucklidge, 1996). Given two tracts p and q with at least two tract points each, the minimum Euclidean distance between a single tract point $\mathbf{p}_i \in p$ and tract q is denoted by:

$$d_{\min}(\mathbf{p}_i, q) = \min_{\mathbf{q}_j \in q} \|\mathbf{p}_i - \mathbf{q}_j\|. \quad (4.10)$$

The one-sided *Hausdorff distance* $d_{\text{HD}}(p, q)$ is then defined as the maximum of the shortest distances:

$$d_{\text{HD}}(p, q) = \max_{\mathbf{p}_i \in p} d_{\min}(\mathbf{p}_i, q). \quad (4.11)$$

Using a shorthand notation, Eqs. 4.10 and 4.11 reduce to

$$d_{\text{HD}}(p, q) = \max_{\mathbf{p}_i \in p} \min_{\mathbf{q}_j \in q} \|\mathbf{p}_i - \mathbf{q}_j\|. \quad (4.12)$$

The one-sided *Hausdorff distance* is not symmetric ($d_{\overline{HD}}(p, q) \neq d_{\overline{HD}}(q, p)$) and does not satisfy the requirements for a similarity measure (see Eq. 2.33 in section 2.6). Hence, instead of the one-sided *Hausdorff distance*, the two-sided or bidirectional *Hausdorff distance* $d_{HD}(p, q)$ is used. It is defined as the maximum of both one-sided *Hausdorff distances* $d_{\overline{HD}}(p, q)$ and $d_{\overline{HD}}(q, p)$:

$$d_{HD}(p, q) = \max(d_{\overline{HD}}(p, q), d_{\overline{HD}}(q, p)). \quad (4.13)$$

4.6 Matched point distance (MPD)

While the *Hausdorff distance* considers all points of the tracts, it neglects their sequential arrangement. In order to take the course of fiber tracts into account, the *matched point distance* (MPD) for two tracts p, q with an arbitrary number of points N, M is defined as the average distance between consecutive, corresponding points of the two tracts³. For the computation of the MPD, three cases are distinguished according to the number of tract points. A pictorial illustration for all cases is shown in Figure 4.2. If the number of tract points is equal ($N = M$), the MPD is the average distance between the consecutive, corresponding points of both tracts (see Figure 4.2a). If $N < M$, a matched, partial tract $\bar{q} = \{q_1, \dots, q_N\}$ is defined that has the same number of tract points as p . Then, the MPD is the sum of the average distance between the first N consecutive tract points of p, q (i.e., $d_{\overline{MPD}}(p, \bar{q})$) and the average distance between the last point of p and the remaining points of q (see Figure 4.2b). If $N > M$, $\bar{p} = \{p_1, \dots, p_M\}$ is the matched, partial tract of p with the same number of points as q . The MPD is then computed in an analogous way as in the previous case with $N < M$ (see Figure 4.2c).

The one sided *matched point distance* $d_{\overline{MPD}}(p, q)$ can now be formally defined:

$$d_{\overline{MPD}}(p, q) = \begin{cases} \frac{1}{S} \sum_{i=1}^N \|p_i - q_i\| & \text{if } N = M, \\ d_{\overline{MPD}}(p, \bar{q}) + \frac{1}{S} \sum_{i=N+1}^M \|p_N - q_i\| & \text{if } N < M, \\ d_{\overline{MPD}}(\bar{p}, q) + \frac{1}{S} \sum_{i=M+1}^N \|p_i - q_M\| & \text{if } N > M, \end{cases} \quad (4.14)$$

where S is $S = \max(N, M)$. To take into account tracts that traverse in opposite directions, the one sided matched point distance $d_{\overline{MPD}}(p, r)$ is computed for p and a tract $r = \{q_M, q_{M-1}, \dots, q_1\}$, for which the order of points of q is reversed. Finally, the *matched point distance* $d_{MPD}(p, q)$ is defined as the minimum between the two one sided *matched point distances* $d_{\overline{MPD}}(p, q)$ and $d_{\overline{MPD}}(p, r)$:

$$d_{MPD}(p, q) = \min(d_{\overline{MPD}}(p, q), d_{\overline{MPD}}(p, r)). \quad (4.15)$$

4.7 Combined distance measure (CD)

The *combined distance* (CD) measure is based on the assumption that the spatial location, the orientation and the shape of the tract are the most important characteristics of fiber tracts.

³The definition of the MPD was inspired by Visser et al. (2011) who also computed the average distance between successive points, but only for tracts that have been resampled to consist of an equal number of tract points.

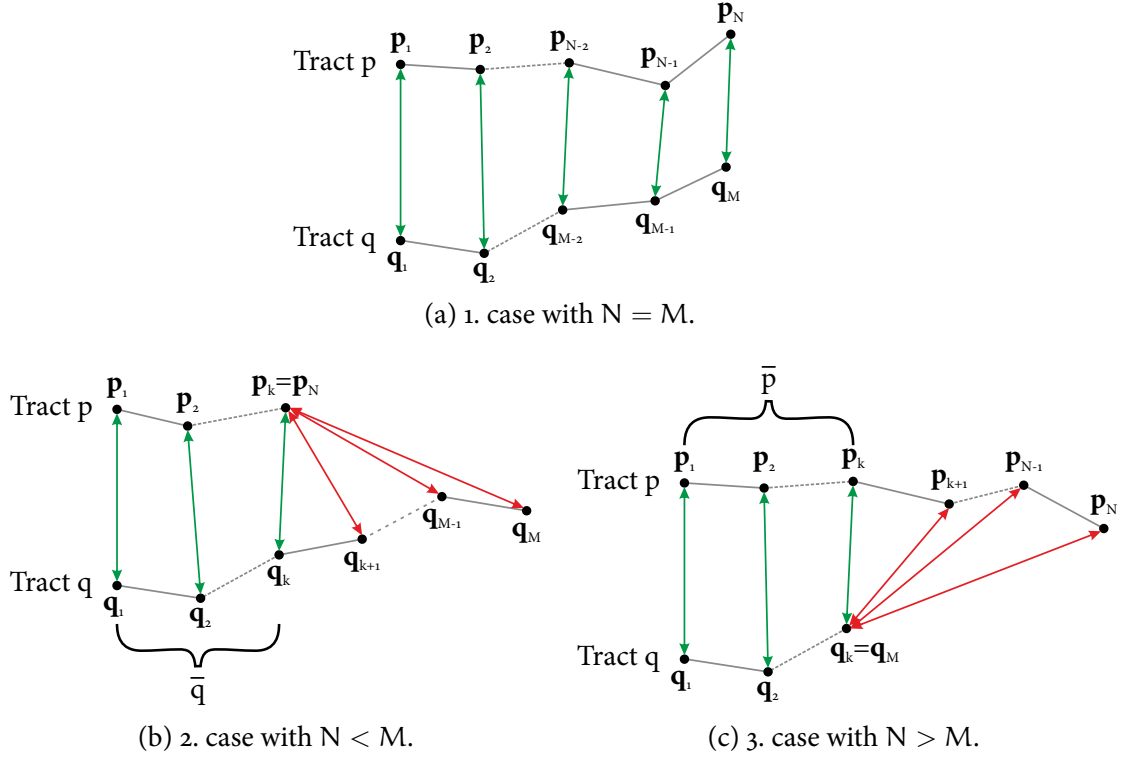


Figure 4.2: Illustration of the *matched point distance* (MPD). For two tracts p and q with N, M tract points, the MPD takes into account the distances between tract points and their successive order. If the number of points N, M is equal for both tracts, the MPD is the average distance between consecutive, corresponding points for the two tracts. This is shown in (a) where the distances that have to be averaged are shown in green. If the number of tract points is different and $N < M$ (b), a matched partial tract \bar{q} is defined for q . The MPD is then the sum of the average distance between the first N consecutive, corresponding tract points of p, q (i.e., $d_{\text{MPD}}(p, \bar{q})$, distances in green) and the average distance between the last point of p and the remaining points of q (distances in red). If $N > M$ (c), a matched partial tract \bar{p} is defined for p and the MPD is computed analogously as in (b).

Consequently, the CD uses all these properties to determine the similarity between two tracts p , q with N , M tract points. To determine the spatial location and the orientation of the tracts, the *tract centroid distance* $d_{TCD}(p, q)$ and the *tract orientation similarity* $d_{TOS}(p, q)$ are used (see section 4.2 and 4.3). To cope with incomplete tracts⁴, partial subtracts $\overline{p'}$ and $\overline{q'}$ are extracted (see below). For these subtracts the *tract centroid distance* $d_{TCD}(\overline{p'}, \overline{q'})$ and the *tract shape similarity* $d_{TSS}(\overline{p'}, \overline{q'})$ (see below) is computed. The one-sided *combined distance measure* $d_{\overline{CD}}(p, q)$ is then defined as the sum of the individual measures⁵ that are normalized between 0 and 1:

$$d_{\overline{CD}}(p, q) = \frac{1}{5} (d_{TCD}(p, q) + d_{TOS}(p, q) + d_{TCD}(\overline{p'}, \overline{q'}) + d_{TSS}(\overline{p'}, \overline{q'})) . \quad (4.16)$$

To account for tracts that traverse in opposite directions, $d_{\overline{CD}}$ is also computed for p and the reversed tract $r = \{q_M, q_{M-1}, \dots, q_1\}$. The similarity $d_{CD}(p, q)$ between tract p and q is then defined as the minimum of $d_{\overline{CD}}(p, q)$ and $d_{\overline{CD}}(p, r)$:

$$d_{CD}(p, q) = \min (d_{\overline{CD}}(p, q), d_{\overline{CD}}(p, r)) . \quad (4.17)$$

While the computation of the *tract centroid distance* $d_{TCD}(p, q)$ and the *tract orientation similarity* $d_{TOS}(p, q)$ has been already described in section 4.2 and 4.3, the extraction of the partial subtracts $\overline{p'}$ and $\overline{q'}$ as well as the computation of the partial tract shape similarity $d_{TSS}(\overline{p'}, \overline{q'})$ is described in the following subsections.

a) Partial tract similarity

To cope with incomplete tracts, a partial tract matching and a shape analysis of the matched tract parts is performed. The shape analysis of the tract parts is based on a method to compare polygonal lines, developed by Attalla and Siy (2005). By using tract resampling and polygonal approximation, tracts are divided into segments before unique properties of the segments are extracted. To assess the similarity between the segments, a segment-based shape descriptor is defined on the basis of the segment properties.

b) Tract resampling

For the computation of the shape similarity, the trajectory of p and q is approximated with newly generated tracts p' , q' that consist of equidistant distributed tract points $p' = \{p'_1, \dots, p'_{N'}\}$ and $q' = \{q'_1, \dots, q'_{M'}\}$, where $p'_1 = p_1$, $p'_{N'} = p_N$, $q'_1 = q_1$, $q'_{M'} = q_M$ (see Figure 4.3a). Intermediate points between $\{p'_1, \dots, p'_{N'}\}$ and $\{q'_1, \dots, q'_{M'}\}$ are scattered along the corresponding trajectories of p , q . Since the level of detail can be adapted by modifying the distance l between the new tract points, polygonal approximation provides flexibility with respect to specific requirements and constraints. By increasing the distance l between the new tract points, the raw number of points is reduced and so is the computational complexity.

⁴Incomplete tracts do not traverse along an entire fiber bundle, but stop prematurely (e.g., in the middle of a bundle).

⁵It might be also beneficial to use specific weights to favor or to penalize the individual measures (Güllmar et al., 2008).

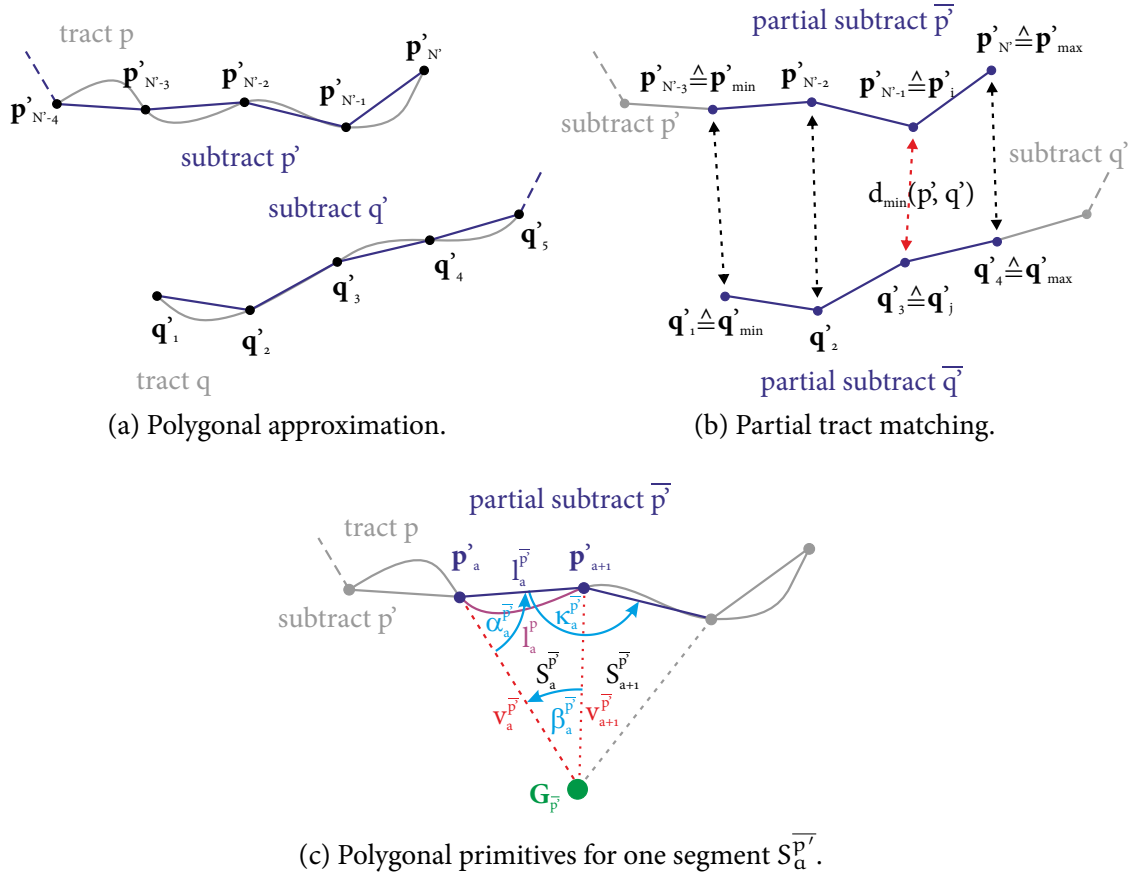


Figure 4.3: Illustration of sub-stages during the generation of the shape similarity for the *combined distance measure* (CD). For the computation of the *shape similarity* for two tracts p and q with N, M tract points, the tract resolution of p and q is reduced and the trajectory is approximated with newly generated tracts p', q' that consist of equidistantly distributed points (a). For p' and q' the closest points p'_i and q'_j are identified (see section 4.4) and used to define corresponding partial subtracts $\overline{p'}$, $\overline{q'}$ (b). Using the tract points of the partial subtract $\overline{p'}$, the partial tract centroid $G_{\overline{p'}}$ is computed and the tract segments $S_a^{\overline{p'}}$ are defined. For each segment $S_a^{\overline{p'}}$, unique properties are identified and employed in subsequent steps to determine the *shape similarity* $d_{TSS}(\overline{p'}, \overline{q'})$ between $\overline{p'}$ and $\overline{q'}$ (c).

c) Partial tract matching

A matching of p' and q' is performed to extract the corresponding, partial subtracts $\overline{p'} \subseteq p'$ and $\overline{q'} \subseteq q'$. For this purpose, the *minimum distance* $d_{MD}(p', q')$ between p' and q' is computed, and the two tract points $p'_i \in p'$, $q'_j \in q'$ that minimize this distance are identified (see section 4.4 *Minimum tract distance*). To extract the corresponding, partial subtracts $\overline{p'}$ and $\overline{q'}$, the matched points p'_i , q'_j are employed (see Figure 4.3b). Formally, $\overline{p'}$ and $\overline{q'}$ are given by:

$$\overline{p'} = \{p'_{\min}, \dots, p'_i, \dots, p'_{\max}\}, \quad (4.18)$$

$$\overline{q'} = \{q'_{\min}, \dots, q'_j, \dots, q'_{\max}\}, \quad (4.19)$$

with:

$$\begin{aligned} p'_{\min} \in \{p'_1, \dots, p'_i\} & \quad \wedge \quad p'_{\max} \in \{p'_i, \dots, p'_{N'}\}, \\ q'_{\min} \in \{q'_1, \dots, q'_j\} & \quad \wedge \quad q'_{\max} \in \{q'_j, \dots, q'_{M'}\}. \end{aligned} \quad (4.20)$$

The four unknown variables (p'_{\min} , p'_{\max} , q'_{\min} , q'_{\max}) are points of p' and q' . They define the starting points, the end points and implicitly the lower and upper parts of $\overline{p'}$ and $\overline{q'}$:

$$\begin{array}{cc} \underbrace{\{p'_{\min}, \dots, p'_i\}}_{\text{lower part of } \overline{p'}} & \underbrace{\{p'_i, \dots, p'_{\max}\}}_{\text{upper part of } \overline{p'}} \\ \underbrace{\{q'_{\min}, \dots, q'_j\}}_{\text{lower part of } \overline{q'}} & \underbrace{\{q'_j, \dots, q'_{\max}\}}_{\text{upper part of } \overline{q'}}. \end{array} \quad (4.21)$$

To qualify as starting and end points, the four unknown variables (Eq. 4.20) have to satisfy the following requirements:

$$|\overline{p'}| = |\overline{q'}|, \quad (4.22)$$

$$|\{p'_{\min}, \dots, p'_i\}| = |\{q'_{\min}, \dots, q'_j\}|, \quad (4.23)$$

$$|\{p'_i, \dots, p'_{\max}\}| = |\{q'_j, \dots, q'_{\max}\}|, \quad (4.24)$$

$$p'_{\min} = p'_1 \quad \vee \quad q'_{\min} = q'_1, \quad (4.25)$$

$$p'_{\max} = p'_{N'} \quad \vee \quad q'_{\max} = q'_{M'}. \quad (4.26)$$

This results in two corresponding, partial subtracts $\overline{p'}$, $\overline{q'}$ having an identical number of tract points (Eq. 4.22), and maximized length (Eqs. 4.25, 4.26). For their lower and upper parts, the number of tract points is also identical (Eqs. 4.23, 4.24).

d) Extraction of partial tract features

In order to accurately describe the shape of the subtracts, a set of partial tract properties (shape descriptors) is extracted. For simplification but without loss of generality, it is assumed that the requirements (Eqs. 4.22–4.26) are fulfilled and $p'_{\min} = p'_1$, $p'_{\max} = p'_{N'}$, $q'_{\min} = q'_1$, $q'_{\max} = q'_{M'}$. Consequently, the subtracts $\overline{p'}$, $\overline{q'}$ are identical to the tracts p' , q' . The individual

segments of $\overline{p'}$ and $\overline{q'}$ are denoted with $S^{\overline{p'}}$, $S^{\overline{q'}}$ and are in each case defined by two successive points p'_a, p'_{a+1} and q'_b, q'_{b+1} :

$$\begin{aligned} S^{\overline{p'}} &= \{S_1^{\overline{p'}}, \dots, S_{N'-1}^{\overline{p'}}\} & \text{with} & & S_a^{\overline{p'}} &= \{p'_a, p'_{a+1}\}, \\ S^{\overline{q'}} &= \{S_1^{\overline{q'}}, \dots, S_{M'-1}^{\overline{q'}}\} & \text{with} & & S_b^{\overline{q'}} &= \{q'_b, q'_{b+1}\}. \end{aligned} \quad (4.27)$$

Along with the centroid $G_{\overline{p'}}$ and the tract points of the segments $S^{\overline{p'}}$, polygonal primitives are extracted. For each segment $S_a^{\overline{p'}} \in S^{\overline{p'}}$, the following primitives are computed (see Figure 4.3c):

- arc length l_a^p of p bounded by p'_a and p'_{a+1} ,
- length $l_a^{\overline{p'}}$ of line $\overline{p'_a p'_{a+1}}$,
- length $v_a^{\overline{p'}}$ of line $\overline{G_{\overline{p'}} p'_a}$,
- angle $\alpha_a^{\overline{p'}}$ between line $\overline{G_{\overline{p'}} p'_a}$ and line $\overline{p'_a p'_{a+1}}$,
- angle $\beta_a^{\overline{p'}}$ between line $\overline{G_{\overline{p'}} p'_a}$ and line $\overline{G_{\overline{p'}} p'_{a+1}}$,
- angle $\kappa_a^{\overline{p'}}$ between neighboring segments $S_a^{\overline{p'}}$ and $S_{a+1}^{\overline{p'}}$,
- smoothness factor $m_a^{\overline{p'}} = l_a^{\overline{p'}} / l_a^p$.

As a result, the sets $v^{\overline{p'}}$, $\alpha^{\overline{p'}}$, $\beta^{\overline{p'}}$, $\kappa^{\overline{p'}}$, $l^{\overline{p'}}$, l^p , and $m^{\overline{p'}}$ are obtained and the primitives of every set are normalized with respect to the maximum value in the set. For subtract $\overline{q'}$ the sets are defined and computed analogously.

e) Elastic segment matching and partial tract *shape & segment orientation similarity*

For two segments $S_a^{\overline{p'}}$ and $S_b^{\overline{q'}}$, the *shape similarity* $d_s(\overline{p'_a}, \overline{q'_b})$ is defined as the normalized sum of the absolute property differences, whilst the *segment orientation similarity* $d_\kappa(\overline{p'_a}, \overline{q'_b})$ is the absolute difference between $\kappa_a^{\overline{p'}}$ and $\kappa_b^{\overline{q'}}$:

$$d_s(S_a^{\overline{p'}}, S_b^{\overline{q'}}) = \frac{1}{4} \left(\begin{aligned} &\text{abs} \left(m_a^{\overline{p'}} - m_b^{\overline{q'}} \right) + \text{abs} \left(v_a^{\overline{p'}} - v_b^{\overline{q'}} \right) + \\ &\text{abs} \left(\alpha_a^{\overline{p'}} - \alpha_b^{\overline{q'}} \right) + \text{abs} \left(\beta_a^{\overline{p'}} - \beta_b^{\overline{q'}} \right) \end{aligned} \right), \quad (4.28)$$

$$d_\kappa(S_a^{\overline{p'}}, S_b^{\overline{q'}}) = \text{abs} \left(\kappa_a^{\overline{p'}} - \kappa_b^{\overline{q'}} \right). \quad (4.29)$$

To compute the *shape similarity* $d_s(S^{\overline{p'}}, S^{\overline{q'}})$ and the *segment orientation similarity* $d_\kappa(S^{\overline{p'}}, S^{\overline{q'}})$ for all segments of the subtracts $\overline{q'}$ and $\overline{q'}$ elastic segment matching is used. This compensates a potential mismatch that may occur during the partial tract matching of the resampled tracts p' and q' . The elastic matching is accomplished by computing the similarity for corresponding

segments of $\overline{p'}$ and $\overline{q'}$ as well as for the k neighboring segments. For $a, b \geq 1$, the *shape* and the *segment orientation similarity* d_s, d_κ are then defined separately:

$$d_s(\overline{S^{p'}}, \overline{S^{q'}}) = \frac{1}{|\overline{S^{p'}}|} \sum_{a=1}^{|\overline{S^{p'}}|} \min_{b=a-k}^{a+k} \left(d_s(\overline{S_a^{p'}}, \overline{S_b^{q'}}) \right), \quad (4.30)$$

$$d_\kappa(\overline{S^{p'}}, \overline{S^{q'}}) = \frac{1}{|\overline{S^{p'}}|} \sum_{a=1}^{|\overline{S^{p'}}|} \min_{b=a-k}^{a+k} \left(d_\kappa(\overline{S_a^{p'}}, \overline{S_b^{q'}}) \right). \quad (4.31)$$

f) Length penalty and tract shape similarity

A segment-wise assessment of tract similarities is only reasonable if an adequate number of segments is available for the computation of $d_s(\overline{p'}, \overline{q'})$ and $d_\kappa(\overline{p'}, \overline{q'})$ (see Eqs. 4.30–4.31). If either $\overline{p'}$ or $\overline{q'}$ consists only of a small subset of p' (q'), the few segments $\overline{S^{p'}}$ ($\overline{S^{q'}}$) are usually not sufficient to accurately determine the segment-wise similarity between $\overline{p'}$ and $\overline{q'}$. In this case, resulting similarities are often characterized by high agreement rates due to few similar segments in $\overline{S^{p'}}$ and $\overline{S^{q'}}$, regardless whether the tracts are resembled well or not. To correct for the adverse effect if the subset P'_s is too small, a correction-factor $C_{\text{corr}}(\overline{p'}, \overline{q'})$ is incorporated as a penalty which is based on the ratio of subtract length and the length of the remaining tract parts.

For tract p' , let $I = \overline{p'}$ be the part of p' that has been matched and $J = p' \setminus \overline{p'}$ the remainder of p' without $\overline{p'}$. The length of I, J is denoted as l^I, l^J and the normalized length as $l_{\text{norm}}^I = l^I / |I|$, $l_{\text{norm}}^J = l^J / |J|$. The penalty for p' is then defined as the ratio between the normalized length of I, J and is given by $C_{\text{corr}}(\overline{p'}) = l_{\text{norm}}^I / l_{\text{norm}}^J$. For $\overline{q'}$, the correction-factor $C_{\text{corr}}(\overline{q'})$ is defined analogously. The correction factor for both subtracts $\overline{p'}, \overline{q'}$ is then denoted with $C_{\text{corr}}(\overline{p'}, \overline{q'})$ and given by:

$$C_{\text{corr}}(\overline{p'}, \overline{q'}) = \begin{cases} C_{\text{corr}}(\overline{p'}) + C_{\text{corr}}(\overline{q'}), & \text{if } C_{\text{corr}}(\overline{p'}) + C_{\text{corr}}(\overline{q'}) > 1, \\ 1, & \text{otherwise.} \end{cases} \quad (4.32)$$

To compensate for few segments, the correction factor $C_{\text{corr}}(\overline{p'}, \overline{q'})$ is applied to the segment-wise similarity measures d_s (Eq. 4.30) and d_κ (Eq. 4.31). Finally, this yields the *tract shape similarity* $d_{\text{TSS}}(\overline{p'}, \overline{q'})$:

$$d_{\text{TSS}}(\overline{p'}, \overline{q'}) = C_{\text{corr}}(\overline{p'}, \overline{q'}) \times \left(d_s(\overline{S^{p'}}, \overline{S^{q'}}) + d_\kappa(\overline{S^{p'}}, \overline{S^{q'}}) \right). \quad (4.33)$$

5

Chapter 5

Technical aspects and implementation details

Nothing is particularly hard if you divide it into small jobs.

— Henry Ford (1863–1947)

The clustering framework was developed to enable fast cluster analysis of large tractography datasets. With the theoretical background for the clustering algorithm CATSER and various similarity measures having been described in the previous chapters 3 and 4, this chapter focuses on particular aspects of the implementation of the clustering framework. A general overview about the implementation of the framework is provided and concepts for the utilization of multiple processing cores for parallel computing are described. A detailed description for the simultaneous processing of similarity measures in a multiprocessing environment is given. Finally, two distinct algorithms for the computation of the Hausdorff distance are presented that are used in the forthcoming chapter to assess the potential of GPU systems for the computation of similarity measures in comparison to conventional CPU systems.

5.1 General implementation details

To facilitate fast processing of large datasets the framework is implemented in C++ for 64-bit Linux (x86-64) architectures. Various software packages, such as the *Boost C++ Libraries* (Karlsson, 2005), the *automatically tuned linear algebra software* (ATLAS) and the *AMD core math library* (ACML), are employed. In addition to these specialized libraries, the framework heavily relies on *symmetric multiprocessing* (SMP) architectures with multiple compute cores and a shared memory model. By these means multithreading with T threads is employed to reduce computation time and to accelerate the processing during stages that are computationally demanding (cmp. section 2.7). For a dataset \mathbb{D} with N tracts three basic processing stages exist:

1. Extraction of a random sample of tracts $\mathbb{D}_{\text{in}} \subseteq \mathbb{D}$ with size N_{in} .

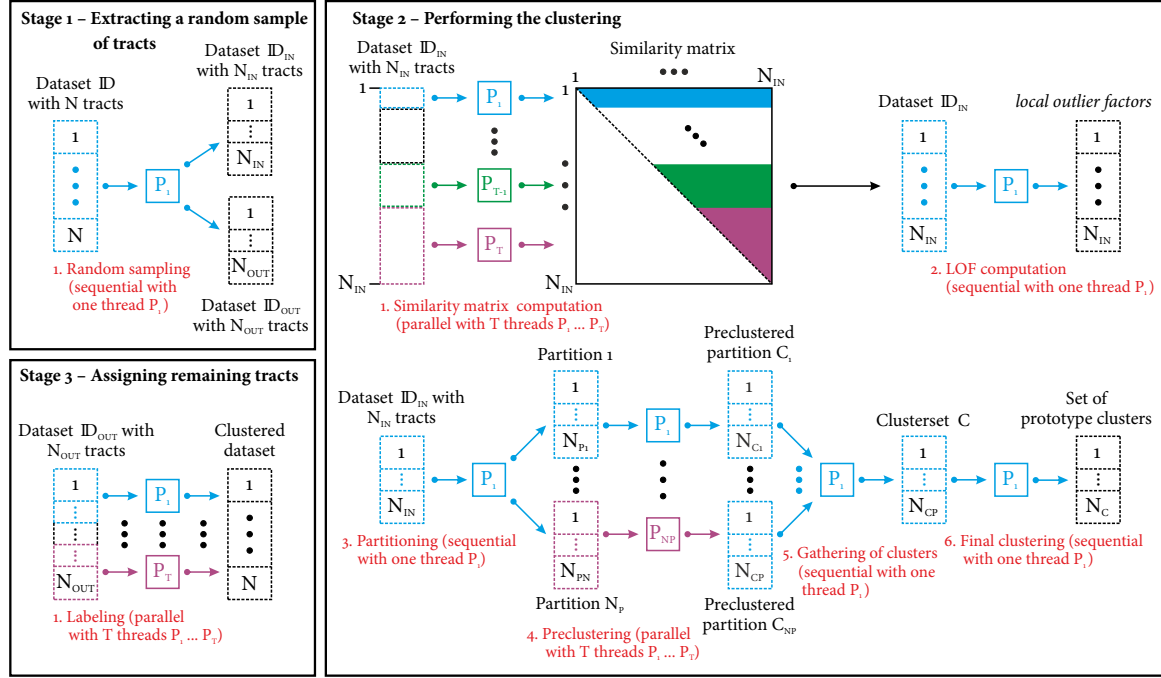


Figure 5.1: Overview of the three fundamental stages of the clustering and their sequential and parallel sub-stages. The figure depicts the way how the data is processed during the clustering stages. It is illustrated, which parts of the clustering are performed either in serial or in parallel and how data is distributed across multiple threads.

2. Cluster analysis for \mathbb{D}_{in} to obtain a set of prototype clusters C with size N_C .
3. Assignment of remaining tracts $\mathbb{D}_{out} = \mathbb{D} \setminus \mathbb{D}_{in}$ to C with size $N_{out} = N - N_{in}$.

Reduction of processing time is accomplished by optimizing and parallelizing computationally demanding parts of these stages. A general overview of both, the sequential and the parallel parts in the framework, is presented in Figure 5.1. A more detailed description is given in the following sections.

5.1.1 Extracting a random sample of tracts (stage 1)

Finding a random sample \mathbb{D}_{in} of tracts in dataset \mathbb{D} is computationally not demanding ($O(n)$ time complexity) and can be carried out serially. Nonetheless, care has to be taken in performing random sampling. Unbiased random sampling is essential to preserve the distribution of tracts as well as the information about the geometry of the bundles. It also prevents the introduction of substantial density variations in the dataset, which is especially important for the calculation of *local outlier factors*.

Random sampling of N_{in} tracts from the finite dataset \mathbb{D} can be regarded as generating a random permutation or performing a random shuffle for all tracts in \mathbb{D} and selecting the first N_{in} tracts in the shuffled set. If the probability for each permutation is equally alike, the

random shuffle is unbiased. Since the *Fisher-Yates shuffle* (Fisher and Yates, 1948; Durstenfeld, 1964) guarantees unbiased results, it is employed for the extraction of the sample.

5.1.2 Performing the clustering (stage 2)

Performing the cluster analysis for the reduced set of tracts \mathbb{D}_{in} consists of several sub-stages – precalculation of the *local outlier factors*, partitioning of the data, preclustering and final clustering that results N_C prototype clusters.

The computationally most expensive operation is the recurrent calculation of pairwise similarities between all tracts in \mathbb{D}_{in} . As the similarities are used multiple times throughout the clustering, computation of the similarities and clustering are separated from each other and similarities are computed only once in the beginning. Since the entirety of similarities has to be available, $(N_{in}^2 - N_{in})/2$ similarities have to be computed for each similarity measure to obtain the triangular $N_{in} \times N_{in}$ similarity matrix \mathcal{S} . This, however, does not cover the computational complexity necessary to compute the pairwise similarity for two tracts. Employing complex similarity measures (e.g., *Hausdorff distance*) increases the computational complexity further. To reduce the computation time, the calculation of the triangular matrix \mathcal{S} is outsourced to multiple threads and performed in parallel. LOFs are then computed by employing the precalculated similarities.

For the preclustering, the *Fisher-Yates shuffle* is employed to shuffle the tracts in \mathbb{D}_{in} , before \mathbb{D}_{in} is partitioned into N_P sets $\{P_1 \dots P_{N_P}\}$ with approximately equal size $\approx N_{in}/N_P$. The partitions are then clustered simultaneously. Thereafter, the final clustering is carried out serially.

5.1.3 Assigning remaining objects (stage 3)

Since the remaining tracts in \mathbb{D}_{out} must be assigned to their most similar prototype clusters of C , the similarity for each tract $p \in \mathbb{D}_{out}$ to all prototype representatives in C has to be computed. This is realized by employing the same similarity measures as in stage 2 (see section 3.2.7).

The N_{out} tracts of \mathbb{D}_{out} are partitioned into T segments of approximately equal size $\approx N_{out}/T$. The segments are assigned to T threads and similarity computations are carried out in parallel.

5.2 Mechanisms for the computation of similarity measures

The most time-consuming parts of the cluster analysis is the arrangement of tracts into clusters and the computation of similarities. As the formation of clusters according to their pairwise similarities (see Figure 5.1, stage 2, steps 3–6) is predominantly a serial task, only minor parts can be computed in parallel. In contrast, the computation of the pairwise similarities between different tracts can efficiently be divided into independent subtasks that are highly suitable for parallel processing¹.

¹Such problems are often termed embarrassingly parallel problems.

Consequently, the framework is designed to utilize multiple processing cores of modern CPU systems. This facilitates the parallel calculation of similarities in order to reduce computation time and to enable fast computation of similarity measures. To prevent redundant operations, multiple similarity measures can be computed once, which enables sharing of results from previously computed features and similarities. To cope with the necessity to integrate additional features and similarity measures into the framework, it is designed to retain simplicity, expandability and performance. New similarity measures can easily be implemented without impairing the capability to parallelize computations. As long as newly integrated similarity measures are in compliance with the framework's specifications, parallelized processing works immediately.

5.2.1 Multithreaded computation of similarity measures

To warrant that the parallelization works efficiently and independent from specific similarity measures, the clustering framework is implemented using task parallelism (see section 2.7 Parallel computing). The similarity computations for the tracts are divided into independent segments and distributed to multiple threads that determine the similarities simultaneously². The similarity computations of each segment are processed by a single thread that computes similarities successively. If multiple similarity measures have to be computed, all similarity measures are computed for the two actual tracts, before the thread advances to compute the similarities of the following tracts. With this approach, threads process the data independently from each other and thread synchronization mechanisms are omitted.

For the computation of similarities two cases are distinguished that are handled slightly differently: the computation of the triangular similarity matrix for generating prototype clusters (cmp. section 2.6 and Figure 5.1 stage 2, step 1) and the computation of similarities for labeling remaining tracts (cmp. section 3.2.7 and Figure 5.1 stage 3, step 1). In practice, the second case can be interpreted as the computation of similarities between two distinct sets of tracts P and Q with N, M tracts³. Similarities are divided into T segments of approximately equal size and processed simultaneously with all T threads in successive order. The concurrent computation for this case is illustrated in Figure 5.2. In the first case similarities are not computed between two distinct sets of tracts but between all tracts of a set of N tracts. Hence, only the triangular similarity matrix has to be computed. Therefore, the matrix is divided into a set of segments with an arbitrary number of contiguous lines that have an approximately equal number of similarity comparisons⁴. Each segment is then processed by one thread in successive order. Figure 5.3 depicts the concurrent processing for this case.

²The computation of specific similarity measures is not parallelized. If only the similarity between two single tracts has to be determined, the processing is performed serially without parallelization.

³P corresponds to the unprocessed tracts and Q to the set of representatives from the prototype clusters.

⁴As the matrix is triangular the number of similarity comparisons differ from line to line.

5.2 Mechanisms for the computation of similarity measures

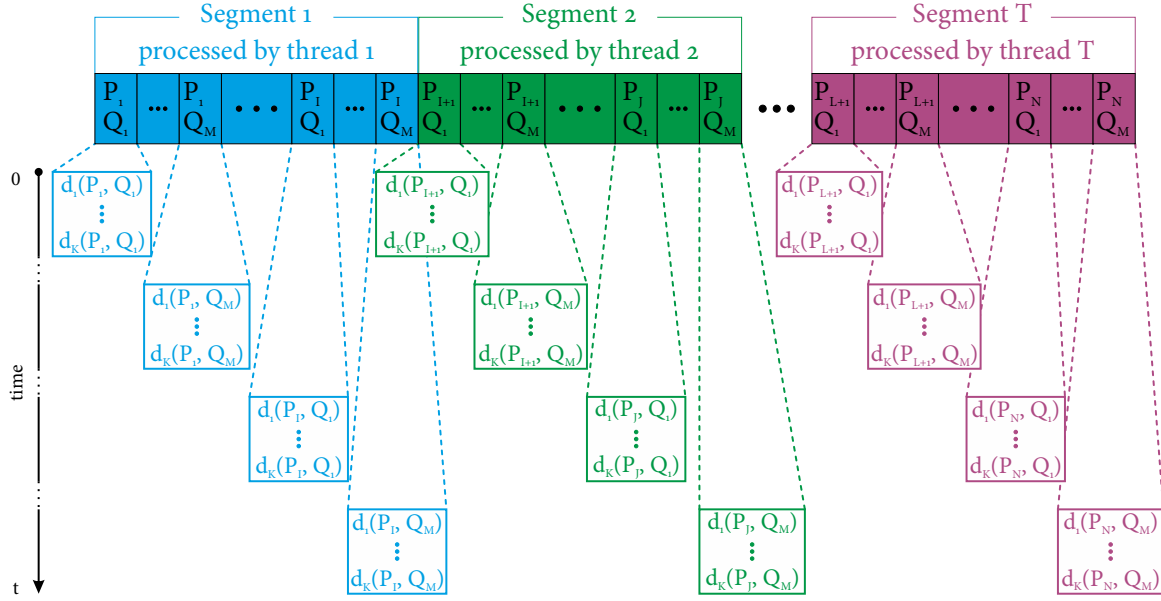


Figure 5.2: Concurrent computation of K similarity measures between two distinct sets of tracts P and Q . Similarity computations are divided into T segments and processed simultaneously with T threads in successive order. All similarity measures (d_1, \dots, d_K) are computed between the two actual tracts before the subsequent tracts are processed.

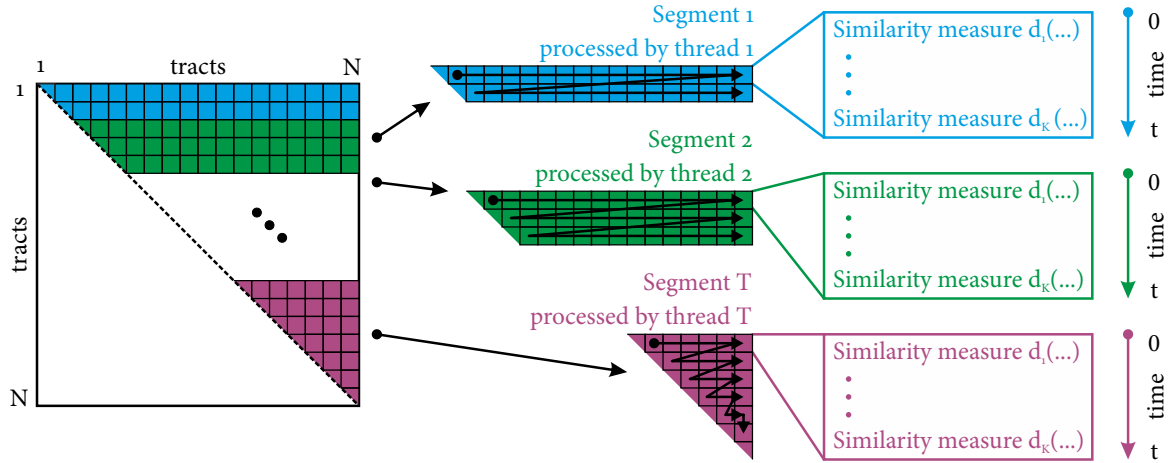


Figure 5.3: Concurrent computation of K similarity measures for a set of tracts. For one set of N tracts, K symmetric similarity matrices are computed in a similar way as illustrated in Figure 5.2. The matrix is divided into T segments of approximately equal size and processed by T threads in parallel. Each thread computes all similarity measures (d_1, \dots, d_K) between the two actual tracts before the subsequent tracts are processed.

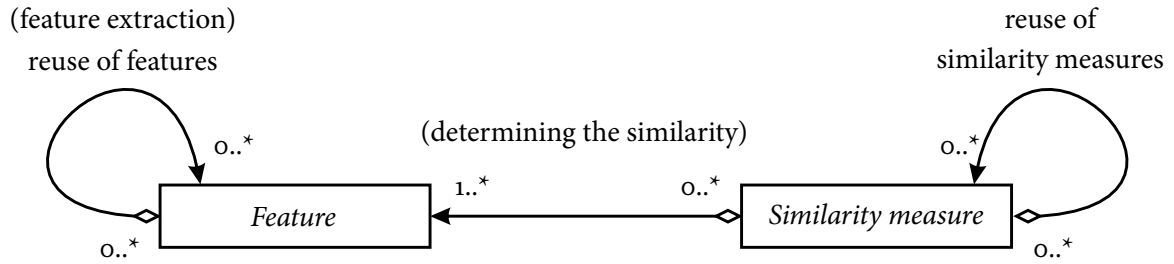


Figure 5.4: Depiction of the extended relationship between features and similarity measures using UML. In order to enable reuseability, features and similarity measures can utilize results from previously computed features and similarity measures.

5.2.2 Reusing results of previously computed features and similarity measures

If multiple features and similarity measures are employed that rely on identical code, additional computational overhead is introduced. In order to keep the computation time to a minimum and preserve computational efficiency, mechanisms are employed to prevent such redundant computations by reusing results of already computed features and similarity measures. To promote reuseability, redundant code is executed only once and results are shared between features and similarity measures. Especially, if identical, computationally intensive operations are deployed to calculate multiple similarity measures, the reuse of previously generated results minimizes computational overhead.

In general, similarity measures rely on features to determine the resemblance of fiber tracts. In the previous definition of features and similarity measures (see section 2.6 and Figure 2.15), the reuseability of features is implicitly declared for the generation of new features by using feature extraction. For similarity measures, however, reuseability of other similarity measures is not defined. By extending the earlier definition (Figure 2.15) to include similarity measures as well, the reuseability of similarity measures becomes possible. The extended relationship between features and similarity measures is illustrated in Figure 5.4 using the notation of the *unified modeling language* – UML (Object Management Group, 2011). Using this definition (Figure 5.4), a similarity measure can now utilize not only features but also additional similarity measures to determine the resemblance between fiber tracts.

In order to highlight the difference between the computation of similarity measures with and without reuse of results, a practical example is given in Figure 5.5 for two similarity measures d_1 and d_2 . In Figure 5.5a both similarity measures implement and utilize the function `complex_operation(...)`, which performs the same, identical operation. As both similarity measures are computed separately, `complex_operation(...)` is executed twice and redundant computations occur. A modified version of the example that employs reuse of results is shown in Figure 5.5b. In this case, `complex_operation(...)` has been extracted from both similarity measures (d_1 , d_2) and implemented as a separate standalone similarity measure d_{comp} . The computation of `complex_operation(...)` is now delegated to d_{comp} . Consequently, d_{comp} is executed only once and the result is relayed to both similarity measures d_1 and d_2 .

5.2 Mechanisms for the computation of similarity measures

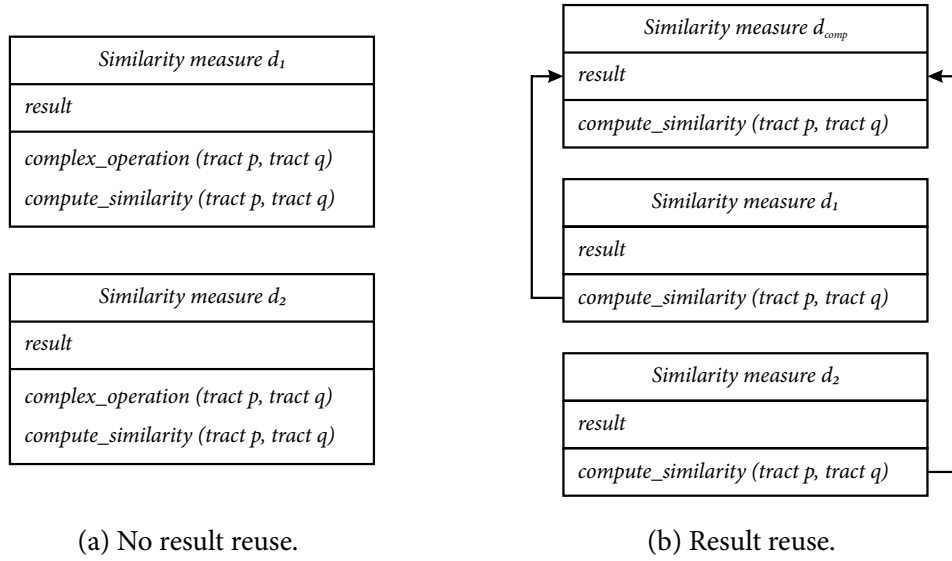


Figure 5.5: Example for the implementation of two similarity measures (d_1 and d_2) with and without result reuse. By computing both similarity measures in (a), the identical function `complex_operation(...)` is executed twice, which results in redundant computations. By implementing `complex_operation(...)` as an independent similarity measure d_{comp} , mechanisms can be used to prevent redundant execution of identical code (b).

5.2.3 Processing and processing order

To perform the computation of multiple features and similarity measures with reusing results, it has to be guaranteed that results are computed before they are used by other features and similarity measures⁵. A valid processing order has to be established by taking into account inter-dependencies between features and similarity measures. Given a set of features and similarity measures without circular dependencies, the order of processing is determined using the algorithm shown in Figure 5.6. For features and similarity measures separate lists for the processing order are generated: the feature processing order P_F and the similarity measure processing order P_{SM} .

During the simultaneous computation of the similarities (see Figures 5.2 and 5.3), all features have to be fully available. Therefore, features are computed in the beginning according to P_F before the parallel processing is initiated. During parallel processing similarity measures are processed in sequential order as specified by P_{SM} .

5.2.4 Memory management

The computation of similarity measures is not only characterized by high processing times but also by high memory usage. If multiple similarity measures are utilized and results are reused, mechanisms have to be established to minimize memory consumption.

⁵In the previous example (see Figure 5.5b), d_{comp} has to be computed before its result is accessed by d_1 or d_2 .

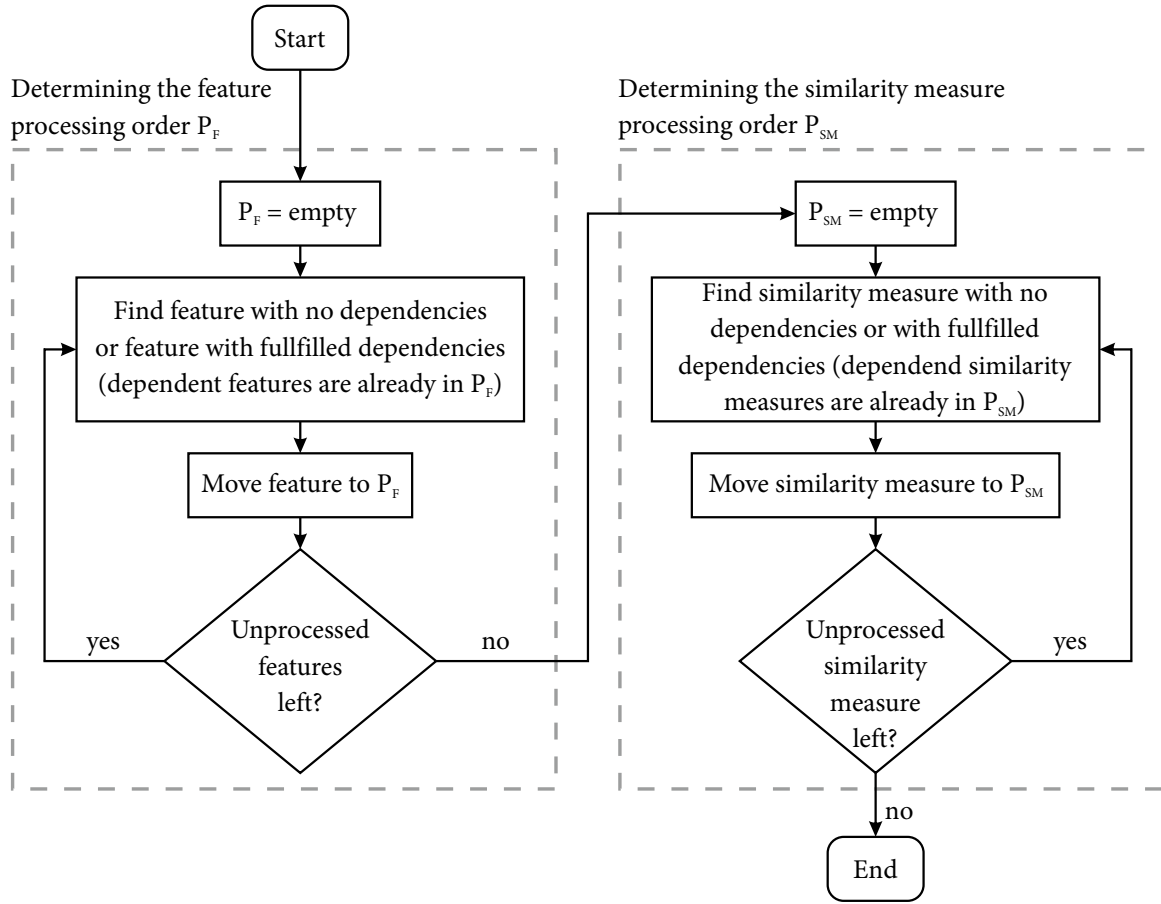


Figure 5.6: Flowchart for determining the feature processing order P_F and similarity measure processing order P_{SM} assuming that circular dependencies are not present.

Using the presented approach for parallel computation of similarity measures (see section 5.2.1), the similarities between two tracts are thoroughly computed before the thread advances to the next pair of tracts. At this point, only those similarities have to be stored in memory that are explicitly requested for the cluster analysis (e.g., by an operator). Intermediate results that are only used by other similarity measures are no longer required and can be discarded. To distinguish between these two types of similarity measures, similarity measures that are explicitly requested by an operator are termed regular similarity measures and similarity measures that are implicitly used by other similarity measures are termed temporary similarity measures⁶. In contrast to similarity measures, features are used multiple times and have to be available throughout the clustering. Features are therefore always categorized as regular features and computed only once in the beginning. Feature results are then shared between those similarity measures that request them.

⁶In the previous example (Figure 5.5b), d_1 and d_2 are regular similarity measures, while d_{comp} is a temporary similarity measure.

5.2 Mechanisms for the computation of similarity measures

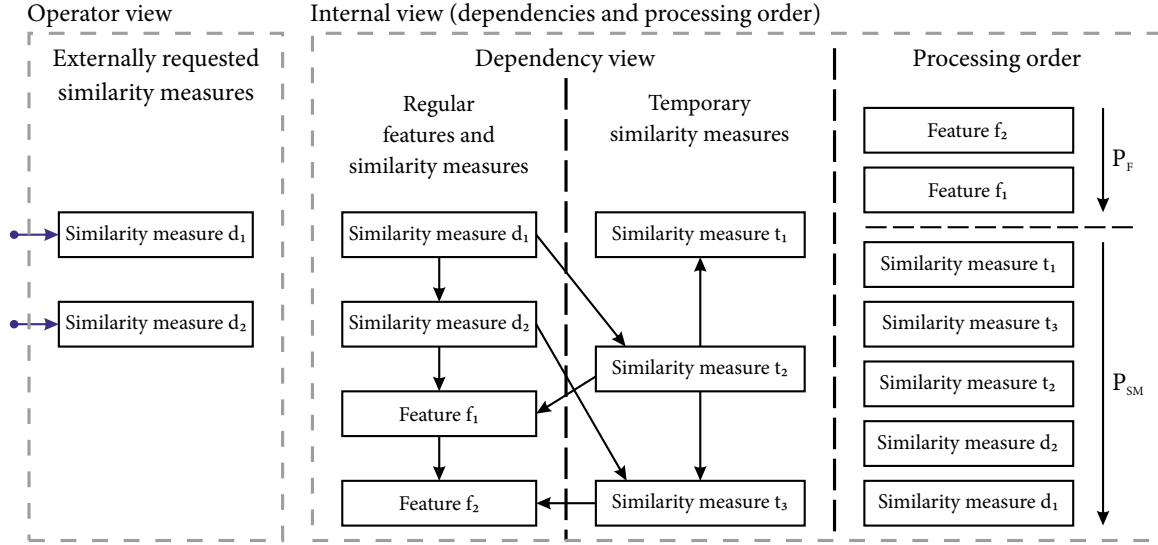


Figure 5.7: Operator view and internal view for two regular similarity measures. The operator view shows the similarity measures d_1 and d_2 that are externally requested by an operator for the cluster analysis. The internal view shows internal mechanisms, including dependencies of related similarity measures, the separation into regular and temporary similarity measures as well as the resulting processing orders P_F and P_{SM} .

Figure 5.7 shows an example for two regular similarity measures d_1 and d_2 (requested by an operator) with various inter-dependencies. An operator view is provided that shows the view of the operator who requested the similarity measures. The corresponding internal view depicts additional internal characteristics that are hidden from the operator. Inter-dependencies between features and similarity measures are shown as well as the separation in regular and temporary features/similarity measures and the resulting processing orders P_F and P_{SM} .

With respect to limited resources and the requirements of a multiprocessing environment, only minimal resources are used to store results in memory. For temporary similarity measures, memory is allocated for each thread to store only similarities for the two actual tracts. If the processing of this set of tracts is finished and the thread advances to process the next pair of tracts, previous results are discarded and allocated memory is used to store the temporary similarities for the new tracts. For regular similarity measures, two different scenarios are considered – the labeling of unprocessed tracts and the computation of the similarity matrices (see section 5.2.1). If similarities are used for the labeling of unprocessed tracts regular similarity measures are used only once (cmp. section 3.2.7 and Figure 5.2). In this case, memory allocation for regular similarity measures is identical to the memory allocation strategy for temporary similarity measures and results can be discarded after they have been processed. The corresponding memory layout for these scenarios is shown in Figure 5.8. If similarity matrices are computed for the clustering and the generation of prototype clusters (cmp. section 3.2.3 and Figure 5.3), results are accessed multiple times and the entire similarity matrices have to be available. Therefore, memory for similarity matrices is allocated once and thread-specific segment boundaries to access the matrices are relayed to each thread. As the utilized similarity measures adhere to certain assumptions (see Eq. 2.33–2.35 in section 2.6),

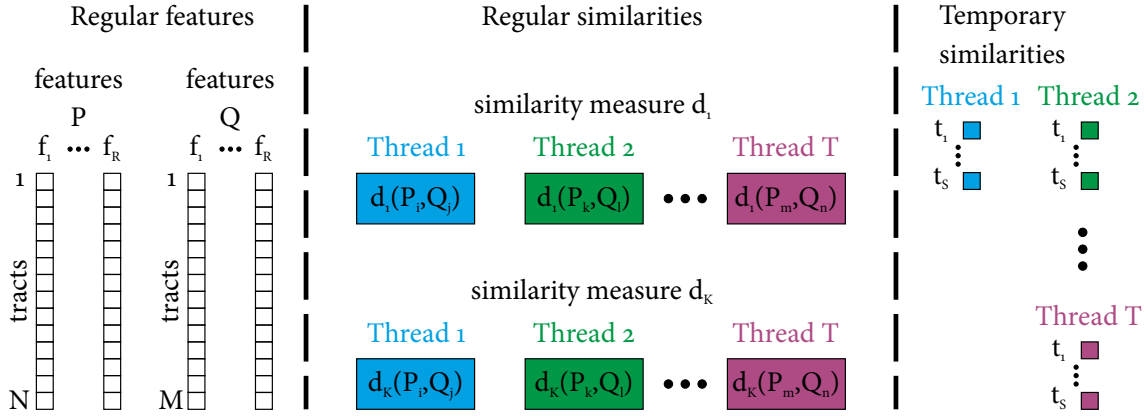


Figure 5.8: Allocated memory that is used to store results from the computation of features and similarities for two set of tracts P and Q with T threads. Reserved memory is shown for R features, K regular similarity measures as well as S temporary similarity measures. Each rectangle corresponds to the memory that is allocated to hold the results of the feature and similarity computations. As results of regular and temporary similarity measures are used only once, only a minimal amount of memory is reserved per thread. During subsequent computations, temporary results are discarded and memory is reused.

the similarity matrix \mathcal{S} of a dataset \mathcal{D} with N elements is triangular and the diagonal elements are zero. To reduce memory consumption, a linear memory layout that stores solely the upper matrix of \mathcal{S} without diagonal elements is used instead of storing the entire $N \times N$ matrix. Matrix \mathcal{S} (cmp. Table 2.1) is thus mapped to memory as followed:

$\mathcal{S}_{\text{lin}}:$	$\mathcal{S}(1, 2)$	\dots	$\mathcal{S}(1, N)$	$\mathcal{S}(2, 3)$	\dots	$\mathcal{S}(2, N)$	\dots	$\mathcal{S}(N-1, N)$
Memory:	0	\dots	$N-1$	N	\dots	$2(N-1)$	\dots	$(N^2 - N)/2$

Using this linear memory model, the value in the similarity matrix at the $(i, j)^{\text{th}}$ position (with $i, j = 1 \dots N$) is given by:

$$\mathcal{S}(i, j) = \begin{cases} \mathcal{S}_{\text{lin}}(N \times i - N + j - 1 - \frac{i^2 + i}{2}) & \text{if } i < j, \\ \mathcal{S}(j, i) & \text{if } i > j, \\ 0 & \text{if } i = j. \end{cases} \quad (5.1)$$

5.3 Algorithms for the computation of the Hausdorff distance

The *Hausdorff distance* – HD (see section 4.5) is employed for various experiments and for the investigation of benefits and disadvantages of *graphics processing units* (GPUs) during the computation of similarity measures (see next chapter). Due to the differing architectures of CPU and GPU systems (cmp. section 2.7), customized implementations have to be used to achieve optimal performance for the different systems. Therefore two algorithms are outlined that are suitable for the computation of the HD with GPU and CPU systems.

5.3 Algorithms for the computation of the Hausdorff distance

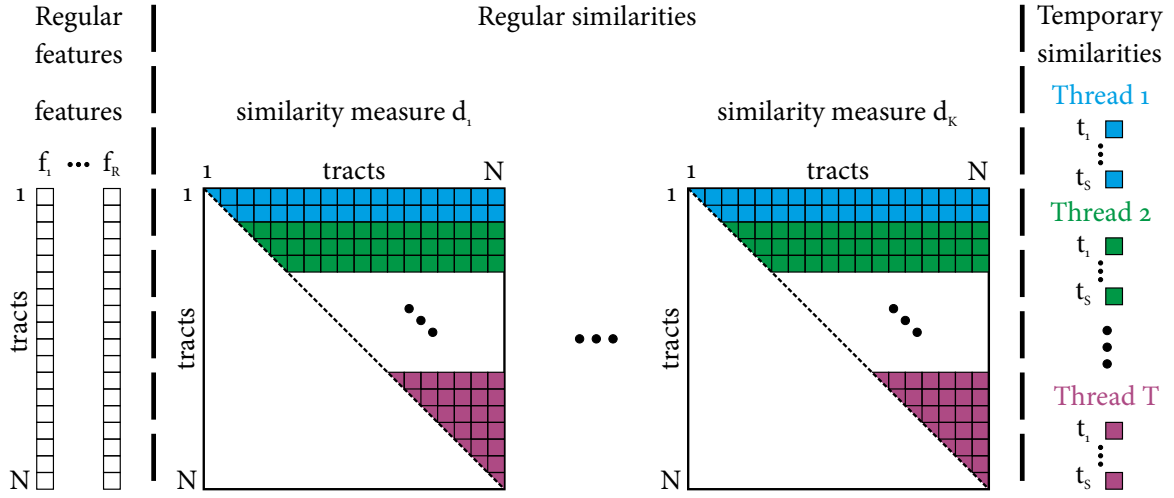


Figure 5.9: Allocated memory that is used to store results from the computation of features and similarity matrices for a set of N tracts with T threads. Reserved memory is shown for R features, K regular similarity measures as well as S temporary similarity measures. Each rectangle corresponds to the memory that is allocated to hold the results of the feature and similarity computations. Temporary similarity measures are used only once. Therefore only a minimal amount of memory is reserved per thread. During subsequent computations, these results are discarded and memory is reused.

The first algorithm is a straightforward implementation of the HD (cmp. section 4.5) that computes $d_{HD}(p, q)$ between two tracts $p = \{p_1, \dots, p_N\}$ and $q = \{q_1, \dots, q_M\}$ with N, M tract points (see Listing 1). At the beginning, the full distance matrix S between all points of p and q is computed. S is then used to determine the one-sided HDs $d_{HD}(p, q)$, $d_{HD}(q, p)$ and finally $d_{HD}(p, q)$. The algorithm contains no diverging paths of execution, which makes it particularly suitable for the processing with GPUs. The computation of S , however, will result in additional, superfluous computations⁷.

The second algorithms is an alternative, more complex implementation that incorporates conditional statements to prevent unneeded operations (see Listing 2). As distances between p_i and all points of q are iteratively assessed to determine $\min(p_i, q)$, it is evaluated in each step if the processing of remaining points of q is still necessary (cmp. section 4.5). Let \min_dist be the minimum distance between p_i and any point of q , determined in an iterative process that computes the distances between p_i and any $q_j \in q$:

$$\min_dist = \min_{q_j \in q} \|p_i - q_j\|. \quad (5.2)$$

The maximum of previously computed minimum distances between $p^s = \{p_1, \dots, p_{i-1}\}$ and q is given by \max_min :

$$\max_min = \max_{p_x \in p^s} \min_dist(p_x, q_j). \quad (5.3)$$

⁷All similarities are computed in advance, regardless of whether they are actually needed or not.

If min_dist falls below max_min for any $\mathbf{q}_j \in q$ during the iterative determination of min_dist (Eq. 5.2), it follows that $\max(\text{min_dist}, \text{max_min}) = \text{max_min}$. As it is guaranteed that $\text{min_dist} < \text{max_min}$ for \mathbf{p}_i and for all $\mathbf{q}_j \in q$, further processing of Eq. 5.2 can be ceased and the next tract point \mathbf{p}_{i+1} can be processed.

Both algorithms (Listing 1 and 2) are integrated into the CPU-based similarity computation framework (cmp. section 5.2). Even though the algorithms themselves are not parallelized for CPU architectures, intrinsic parallelization is achieved by the clustering framework if multiple similarities are computed (cmp. section 5.2). Due to the distinct processing concepts of CPUs and GPUs and the limitation of GPUs, the clustering framework as presented in section 5.2 cannot be transformed to GPU architectures. Hence, only the first algorithm (Listing 1) is specifically implemented for GPUs.

Listings 1: Algorithm for the computation of the *Hausdorff distance*.

Data: Tracts $p = \{\mathbf{p}_1, \dots, \mathbf{p}_N\}$ and $q = \{\mathbf{q}_1, \dots, \mathbf{q}_M\}$ with N and M tract points.

Result: The *Hausdorff distance* $d_{\text{HD}}(p, q)$.

```
// compute  $N \times M$  distance matrix  $S$  between all points of  $p$  and  $q$ 
for  $i = 1$  to  $N$  do
    for  $j = 1$  to  $M$  do
         $S(i, j) = \|\mathbf{p}_i - \mathbf{q}_j\|;$ 

// determine the one sided Hausdorff distance  $d_{\text{HD}}(p, q)$ 
for  $i = 1$  to  $N$  do      // for each point in  $p$ , find minimum distance to  $q$ 
     $m(i) = \text{find\_minimum}(\{D(i, 1), \dots, D(i, M)\});$ 
 $d_{\text{HD}}(p, q) = \text{find\_maximum}(m);$ 

// determine the one sided Hausdorff distance  $d_{\text{HD}}(q, p)$ 
for  $i = 1$  to  $M$  do      // for each point in  $q$ , find minimum distance to  $p$ 
     $m(i) = \text{find\_minimum}(\{D(1, i), \dots, D(N, i)\});$ 
 $d_{\text{HD}}(q, p) = \text{find\_maximum}(m);$ 

 $d_{\text{HD}}(p, q) = \max(d_{\text{HD}}(p, q), d_{\text{HD}}(q, p));$ 
```

Listings 2: Optimized algorithm for the computation of the *Hausdorff distance*.

Data: Tracts $p = \{p_1, \dots, p_N\}$ and $q = \{q_1, \dots, q_M\}$ with N and M tract points.

Result: The *Hausdorff distance* $d_{HD}(p, q)$.

```
// compute one sided Hausdorff distance max_min between p and q
min_dist = inf;
// initialize min_dist by finding the minimum distance between p1 and
  q
for j = 1 to M do
  | if  $\|p_1 - q_j\| < \text{min\_dist}$  then min_dist =  $\|p_1 - q_j\|$ ;
max_min = min_dist;
for i = 2 to N do
  | min_dist = inf;
  | j = 1;
  | // process all points in q and iterate as long as
  |   min_dist > max_min
  | while j < M and min_dist > max_min do
  | | if max_min < min_dist then min_dist =  $\|p_i - q_j\|$ ;
  | | j = j + 1;
  | max_min = max(min_dist, max_min);

// At this point, max_min corresponds to the one sided Hausdorff
  distance  $d_{HD}(p, q)$ 

// now compute max( $d_{HD}(p, q)$ ,  $d_{HD}(q, p)$ )
for i = 1 to M do
  | min_dist = inf;
  | j = 1;
  | // process all points in p and iterate as long as
  |   min_dist > max_min
  | while j < N and min_dist > max_min do
  | | if  $d(p_j, q_i) < \text{min\_dist}$  then min_dist =  $\|p_j - q_i\|$ ;
  | | j = j + 1;
  | max_min = max(min_dist, max_min);

 $d_{HD}(p, q) = \text{max\_min}$ ;
```

6

Chapter 6

Clustering experiments

*The true method of knowledge is
experiment.*

— William Blake (1757–1827)

The experimental investigation of the cluster analysis framework is important to study the properties of the implemented algorithms and to assess the quality of the clustering. For this purpose, extensive experiments were conducted that are presented in this chapter.

Several datasets of healthy subjects were acquired. These datasets were not only used during the clustering experiments, but also to create a voxel-based white matter atlas which was used for the atlas-guided clustering and for the evaluation of results. In order to assess the quality of the clustering quantitatively, the spatial agreement between obtained clusters and the atlas classes was determined. To investigate the computational efficiency of the clustering framework in multiprocessing environments, multiple experiments were conducted.

6.1 Data acquisition and data processing

46 healthy subjects (24 female, 29 ± 8.6 years old; 22 male, 29 ± 10.0 years old) were included and measured on a clinical 3 T whole-body MR-Scanner (Tim Trio, Siemens Healthcare, Erlangen, Germany). The study was approved by the Ethics Committee of the Jena University Hospital. Participants provided informed written consent to participate in this study in accordance to the statement of the Ethics Committee.

For the diffusion tensor acquisition, a conventional single shot twice refocused *echo planar imaging* (EPI) sequence was used with four bipolar diffusion gradients to compensate for eddy currents (Heid, 2000). Fat suppression was achieved by applying a *spectral attenuated inversion recovery* (SPAIR) pulse. A 12 channel phased array matrix head coil was employed for signal reception. To minimize subject motion, special pads were used that secured a tight fit of the heads in the coil. The following acquisition parameters were used: a matrix of 96×96 , 55 slices with a thickness of 2.5 mm each, voxel size = $2.5 \times 2.5 \times 2.5$ mm³, $T_E = 91$ ms, $T_R = 6800$ ms, $\alpha = 90^\circ$ and data sampling acceleration factor 2. Six b_0 images without diffusion weighting (b

= 0 s/mm²) as well as 70 diffusion weighted images sampled with different gradient directions at $b = 1000$ s/mm² were acquired.

Following data acquisition, in-plane interpolation was performed on the MR-Scanner, resulting in an effective voxel size of $1.25 \times 1.25 \times 2.5$ mm³. Due to the head fixation, subject motion was not observed and a retrospective motion correction was not performed. The Diffusion Toolkit (Wang et al., 2007) was used to perform DTI-based analysis and whole brain fiber tractography using the interpolated streamline method (Conturo et al., 1999) with a fixed step-length of 0.5 mm and an angle threshold of 41°. For fiber tracking, three randomly located seed points were placed at intravoxel positions in each voxel of the brain's white matter. The white matter seed mask was derived from the FA maps with a manually-tuned minimum FA threshold of 0.2. Tracts having a length less than 30 mm were subsequently removed from the dataset. On average, each dataset consisted of about 280 000 tracts with more than 20 million tract points per dataset. Tracts were not resampled and differed in both length and number of tract points.

All datasets were spatially normalized in a two step procedure by using the *advanced normalization tools* (ANTs) (Avants et al., 2011). In an initial step, a rigid transformation was applied to the FA maps to register and coarsely align all data to the common FA template (FMRIB58) of the *FMRIB's software library* (FSL) (Jenkinson et al., 2012). A subject-specific template was created by normalizing the datasets with the non-linear template generation framework of ANTs. As a starting point for this non-linear, spatial normalization process, the rigidly transformed FA maps were averaged to produce an initial FA template. The template was refined and improved in four iterations using the *greedy SyN* transformation model and the cross correlation metric of ANTs. The resulting transformation matrices and displacement fields were finally employed to transfer the fiber tracts into the space of the newly generated template. To enable the spatial transformation of fiber tracts, the open source ANTs framework was extended.

6.2 Creation of a white matter atlas

A probabilistic white matter atlas was semi-automatically constructed. Out of the 46 spatially normalized datasets (see previous section), 15 datasets were randomly selected to generate the atlas. 16 white matter structures (bundles) were included in the atlas: forceps major (Fmaj), the frontal projection of the corpus callosum (forceps minor – Fmin) as well as the following bundles of the left and right hemisphere: anterior thalamic radiation (ATR), gyrus part of the cingulum (CGC), hippocampal part of the cingulum (CGH), corticospinal tract (CST), inferior fronto-occipital fasciculus (IFO), temporal part of the superior longitudinal fasciculus (SLFt), uncinate fasciculus (UNC). To delineate these bundles, a set of *regions of interest* (ROIs) was drawn for each bundle, taking into account the guidelines for reproducible extraction by Wakana et al. (2007). For each dataset, tracts that crossed these ROIs were extracted and assigned to the corresponding WM bundle. While this is an efficient and fast way to extract the WM fiber bundles, it only extracts the major parts of the bundles. Tracts that belong to the bundle but do not cross all ROIs are not assigned to the bundle. This probably resulted in a loss of minuscule details for the bundles.

The probabilistic white matter atlas was then created by using the extracted bundles ($K = 16$) of all datasets ($N = 15$). With all these bundles, the 16 prototype classes of the white matter atlas were generated. Each class in the atlas contains all voxels that are associated with the corresponding atlas class and describes how reliably each voxel can be associated with this class. Let A_1, \dots, A_K be the prototype classes. If A is one of these classes, it formally consists of a list of voxels $V^A = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ with an unknown number of voxels n that belong to it. Each $\mathbf{v}_i \in V^A$ is a set of coordinates $\mathbf{v}_i = (v_i^x, v_i^y, v_i^z)$ that describes the position of voxel \mathbf{v}_i in the 3D dataset. The probability that voxel \mathbf{v}_i belongs to class A is denoted by $\phi_A(\mathbf{v}_i)$. As V^A consists only of voxels that belong to A , the probability for each voxel $\mathbf{v}_i \in V^A$ is $\phi_A(\mathbf{v}_i) > 0$. $\phi_A(\mathbf{v}_i)$ is therefore bounded by $]0, 1]$.

To generate the probabilistic atlas a two step procedure was used. During the first step, the probabilities for each fiber bundle were computed individually for each dataset ($N = 15$), before these probabilities were used to generate the final prototype classes in the second step.

For the first step, the computation of the dataset probability ϕ^D is performed individually for each dataset $D \in \{D_1, \dots, D_N\}$. Initially, for each fiber bundle A of dataset D , the tract density ρ_A is determined. Hereby each tract that belongs to bundle A is rasterized to a user-defined 3D grid and all voxels $V^A = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ that contain tracts of A are identified. The tract density $\rho_A(\mathbf{v}_i)$ for voxel \mathbf{v}_i is computed by counting the number of tracts that run through voxel \mathbf{v}_i . To obtain the dataset probability $\phi_A^D(\mathbf{v}_i)$ for voxel \mathbf{v}_i of bundle A , the ratio between the tract density $\rho_A(\mathbf{v}_i)$ and the number of all tracts that occupy voxel \mathbf{v}_i is computed:

$$\phi_A^D(\mathbf{v}_i) = \frac{\rho_A(\mathbf{v}_i)}{\sum_{j=1}^K \rho_j(\mathbf{v}_i)}. \quad (6.1)$$

If only tracts of A occupy voxel \mathbf{v}_i , the probability $\phi_A^D(\mathbf{v}_i)$ is 1.

After computing the dataset probability for each dataset and each fiber bundle the final prototype classes are generated in the second step. For prototype class A , the prototype probability $\phi_A(\mathbf{v}_i)$ in voxel \mathbf{v}_i is defined as the average of all dataset probabilities $\phi_A^j(\mathbf{v}_i)$ for A in voxel \mathbf{v}_i with $j = 1, \dots, N$:

$$\phi_A(\mathbf{v}_i) = \frac{1}{N} \sum_{j=1}^N \phi_A^j(\mathbf{v}_i). \quad (6.2)$$

If there is no voxel in a bundle A to which the corresponding bundles of all N datasets contribute, the maximum bundle probability $\max(\phi_A)$ will be less than 1. To prevent such a degradation of probabilities, the probabilities in each bundle are normalized to a maximum value of 1.

During this prototype generation stage, tracts are rasterized to a 3D grid. As tracts are a set of real-valued points in 3D space, the atlas can be reconstructed for arbitrary grid resolutions. For the experiments, the atlas was reconstructed for an isotropic resolution of 1 mm^3 . Unreliable voxels with probabilities less than 0.3 were removed from the atlas before further processing was performed. An example for a class of the probabilistic atlas is given in Figure 6.1. Volume renderings for a selection of fiber bundles that are defined in the atlas are shown in Figure 6.2. The bundles are overlaid onto the FA volume of a single subject.

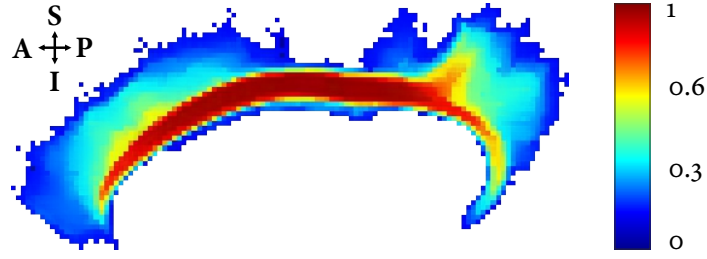


Figure 6.1: Example for an atlas class in the probabilistic white matter atlas. The maximum intensity projection of the probabilities is shown as a pseudo color image for the gyrus part of the left cingulum (CGC_{left}). Regions with high probability ($\phi_{CGC_{left}} = 1$) are colored in red, while regions with low probability ($\phi_{CGC_{left}} = 0$) are shown in blue.

6.3 Spatial matching of clusters and atlas classes

After the clustering of a dataset, the obtained clusters are not ordered and a spatial matching is performed to relate the clusters to their best matching atlas class. For this purpose, the clusters of a dataset \mathbb{D} are rasterized to a 3D grid with the same spatial resolution as the atlas. The tract density ρ_C is computed for all clusters $C \in \{C_1, \dots, C_n\}$ and normalized to a maximum density of 1 for each cluster.

The spatial agreement between all atlas classes and all obtained clusters is then determined. A matching value $\zeta(A, C)$ for an atlas class A and a cluster C is computed that reflects the spatial resemblance between A and C (see below). The larger the value $\zeta(A, C)$ the higher the spatial agreement between A and C . After the computation of the spatial matching value for all atlas classes and all obtained clusters, clusters are iteratively assigned to the best matching atlas class until clusters have been associated with an atlas class. During this process, a one to one mapping is enforced and clusters cannot be assigned to more than one atlas class.

To determine the matching value $\zeta(A, C)$, the intersecting voxels $V^{AC} = \{v_1, \dots, v_M\}$ as well as the non-intersecting voxels $V^A = \{v_1^A, \dots, v_K^A\}$ for A and $V^C = \{v_1^C, \dots, v_L^C\}$ for C are identified. For A and C , the number of all unique voxels ($A \vee C$) is denoted by N with $N = K + L - M$. In the following it is assumed that V^A and V^C are not empty ($K \geq 1 \wedge L \geq 1$) and at least one intersecting voxel exists ($M \geq 1$). Otherwise, the matching value $\zeta(A, C)$ is zero. To assess the similarity in the overlapping regions of A and C , the average of the differences in the intersecting voxels is computed by:

$$\Delta(A, C) = \frac{1}{M} \sum_{i=1}^M (1 - \text{abs}(\rho_A(v_i) - \rho_C(v_i))). \quad (6.3)$$

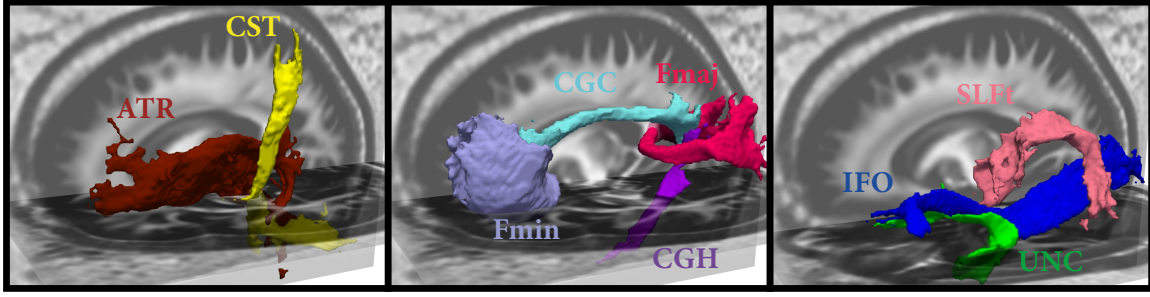


Figure 6.2: This figure shows the volume renderings for a selection of fiber bundles defined in the white matter atlas. From left to right the bundles are: anterior thalamic radiation (ATR_{left}), corticospinal tract (CST_{left}), the forceps minor (F_{min}), gyrus part of the right cingulum (CGC_{right}), hippocampal part of the right cingulum (CGH_{right}), forceps major (F_{maj}), inferior fronto-occipital fasciculus (IFO_{left}), temporal part of the superior longitudinal fasciculus (SLF_{left}), uncinate fasciculus (UNC_{left}).

To take into account the average tract density of non-intersecting voxels for both A and C , $\Delta(A)$ and $\Delta(C)$ were also computed:

$$\Delta(A) = \frac{1}{K} \sum_{i=1}^K (1 - \rho_A(\mathbf{v}_i^A)), \quad (6.4)$$

$$\Delta(C) = \frac{1}{L} \sum_{i=1}^L (1 - \rho_C(\mathbf{v}_i^C)). \quad (6.5)$$

The matching value $\zeta(A, C)$ is then defined as the product of Eqs. 6.3–6.5, weighted by the ratio between the number of intersecting voxels M and the number of all unique voxels N :

$$\zeta(A, C) = \frac{M}{N} \times \Delta(A, C) \times \Delta(A) \times \Delta(C). \quad (6.6)$$

6.4 Experiments

6.4.1 Cluster analysis

For each of the 46 whole brain tractography datasets, the cluster analysis was performed in the atlas space with three different techniques: atlas-guided clustering with CATSER, conventional CATSER clustering (without atlas) as well as standard *hierarchical agglomerative clustering* (HAC) using Ward's method (Ward, 1963). For both atlas-guided and non-atlas-guided clustering with CATSER, the same parameters were used (see below). For HAC, the datasets were far too large to be processed within reasonable time. Therefore, 10 000 tracts were selected randomly from each dataset. The samples were clustered individually and 250 clusters were reconstructed for each dataset.

For the cluster analysis with CATSER, identical parameters empirically derived from prior experiments were used. A sampling size of 10 000 tracts was chosen (Figure 3.2, step 1). The sample was divided into 3 partitions (Figure 3.2, step 3) and clustered during the preclustering

stage (Figure 3.2, step 4). After finishing 80 % of the preclustering iterations (see section 3.2.4), outlier elimination was performed and very small clusters with less than three tracts were removed. During the final clustering stage (Figure 3.2, step 6) the second outlier elimination stage was performed after finishing 60 % of the clustering iterations. This time, clusters with less than 4 tracts were removed. For the computation of the LOFs, the number of neighboring points k was set to 15. The number of representatives was set to a maximum value of 40 for a cluster size of 120 tracts. With respect to Eq. 3.8 this yields $\chi_1 = 120$ and $s = 40/120$. The factor f_r for reassignment of outliers was set to 1 and for the labeling of unprocessed tracts to $f_l = 1.5$ (see section 3.2.7). 250 fiber bundles were reconstructed for each dataset.

For the atlas-guided clustering with CATSER, the white matter atlas was used (see section 6.2). A subsequent step was employed, in which the spatial agreement of the clusters to all atlas bundles was assessed, and clusters were merged if this merging would lead to an increased spatial agreement between the newly formed clusters and the best matching atlas regions. The purpose of this additional step is to guarantee that clusters are not splitted but are completely formed.

To investigate the performance of the similarity measures for multiple subjects and the different clustering techniques, the clustering of all datasets was performed thrice, each time with the three clustering techniques (see above) but different similarity measures. For this experiment, the *combined distance measure* (CD), the *Hausdorff distance* (HD) and the *matched point distance* (MPD) were used (see chapter 4). For CD, the tract approximation distance l was set to 15 mm (see section 4.7). For MPD, tracts were resampled to consist of 10 tract points (see section 4.6). In total, cluster analysis was performed 414 times ($46 \text{ datasets} \times 3 \text{ clustering techniques} \times 3 \text{ similarity measures}$).

With spatial matching (see section 6.3), clusters were identified and related to their corresponding and best matching class in the atlas. To evaluate the quality of the final results, the spatial agreement (see Eq. 3.12) between the rasterized clusters and their corresponding, best matching atlas classes was computed.

6.4.2 Outlier elimination

To demonstrate the benefits of outlier elimination, the effects of different outlier elimination strategies and varying levels of noise on the clustering results were investigated. For this purpose, one dataset that resided in its native space was segmented according to the guidelines by Wakana et al. (2007). The same 16 fiber bundles that are also defined in the atlas (see Figure 6.2) were extracted. For this segmented dataset, unsupervised clustering (without white matter atlas) was performed with varying levels of noise and different sets of outlier elimination parameters. As the correctness of fiber tract clusters are visually hard to grasp due to their inherent complexity, the Euclidean norm between the tract centroids was used as a similarity measure for this clustering experiment (see section 4.2). Contrary to fiber tracts, the distance between tract centroids can be easily depicted in 3D Euclidean space, which allows good visual delineation of the clusters and their shapes. For this experiment, the cluster analysis was performed for three different outlier elimination parameter sets (low, moderate and high outlier elimination, see Table 6.1 for details). In addition, artificial white

noise was added on top of the tract centroids and gradually increased (0 %, 33 %, 66 % and 99 % additional noise).

Table 6.1: Parameters for the three different outlier elimination strategies.

Outlier elimination strategy	Preclustering		Final clustering	
	time point t_1 in %	critical size h_1	time point t_2 in %	critical size h_2
low	95	1	85	4
moderate	80	2	85	6
high	80	4	85	8

The table shows the outlier elimination parameters for the three outlier elimination strategies (low, moderate, high). The outlier elimination is performed during the preclustering (Figure 3.2, step 4) as well as during the final clustering (Figure 3.2, step 6). Clusters that contain no more tracts than the critical cluster size h_1 (h_2) after $t_1\%$ ($t_2\%$) of the clustering has been finished are considered outliers and are removed from the subsequent clustering.

6.4.3 Performance analysis

In order to assess the performance of the clustering in a multiprocessing environment, a performance analysis was conducted using an eight core *symmetric multiprocessing* (SMP) system (CPU system 1, Table 6.2). Effects on the execution time T_P and the relative speedup S_P were investigated by gradually increasing the number of utilized cores P (cmp. section 2.7).

The performance analysis was conducted for unsupervised CATSER clustering without white matter atlas using the parameters stated in section 6.4.1. One dataset \mathbb{D} with 100 000 fiber tracts was used, whereas the reduced random sample consisted of 10 000 tracts. The *Hausdorff distance* (HD), the *combined distance* (CD) and the *matched point distance* (MPD) were used as similarity measures. To impede statistical fluctuations due to running background processes, all computations were repeated 10 times.

The analysis of the clustering framework was divided into three individual parts to identify the parts of the clustering that are suspected to be computationally most critical (cmp. Figure 3.2), as well as to identify the parts that will profit the most from adding additional cores:

- Part 1: Computation of the similarity measures for \mathbb{D}_{in} (Figure 3.2, step 2)
- Part 2: Clustering of the sample dataset \mathbb{D}_{in} (Figure 3.2, steps 3–7):
The performance during the formation of clusters was investigated by gradually increasing the number of parallel clustered partitions.

Table 6.2: Technical details for the employed CPU and GPU systems.

Technical Details	CPU systems		GPU systems		
	System 1	System 2	NVIDIA Quadro FX 1800	NVIDIA Tesla C1060	NVIDIA Tesla C2050
Manufacturer	Intel	AMD	NVIDIA	NVIDIA	NVIDIA
Processor	Xeon L5430	Opteron 8350	G94	GT200	GF100
Clock rate	2.66 GHz	2.0 GHz	1.38 GHz	1.3 GHz	1.15 GHz
Multiprocessor	2	8	8	30	14
Cores/MP	4	4	8	8	32
Cores (total)	8	32	64	240	448
RAM	16 GB	128 GB	2 GB	4 GB	2.6 GB

- Part 3: Labeling of remaining tracts $\mathbb{D}_{\text{out}} = \mathbb{D} \setminus \mathbb{D}_{\text{in}}$ (Figure 3.2, step 8):
By employing identical clustering parameters as in part 2, the performance of the labeling was analyzed.

To investigate the potential of *graphics processing units* – GPUs (section 2.7) in speeding up fiber tract similarity computations, three different GPU systems were employed. The three systems were used to compute the *Hausdorff distance* (see section 5.3, algorithm 1) and to assess the performance. In addition, a 32 core CPU system (CPU system 2, Table 6.2) was employed as gold standard and the performance of the two HD algorithms was assessed (see section 5.3, algorithm 1 and 2). Finally, a comparison between the CPU and GPU systems was conducted. Technical details for the different computing systems are given in Table 6.2.

6.5 Results

6.5.1 Cluster analysis

The clustering of all 46 datasets was successfully performed using the three clustering techniques (atlas-guided CATSER, CATSER, HAC) and all three similarity measures (CD, HD, MPD). For each clustering experiment and each dataset, the clusters were matched to the atlas classes and the best matching cluster was determined for each class. Clusters for one exemplary dataset, processed with the atlas-guided CATSER clustering and the *combined distance measure* (CD), are shown in Figure 6.3. All extracted bundles of the dataset are displayed in the top row. To enhance the visualization, the bundles in the upper row have been divided into three groups: bundles of the left hemisphere (left image), bundles that cross both hemispheres (middle image) and bundles of the right hemisphere (right image). Fiber bundles are displayed in distinct colors, and tracts that belong to the same cluster are

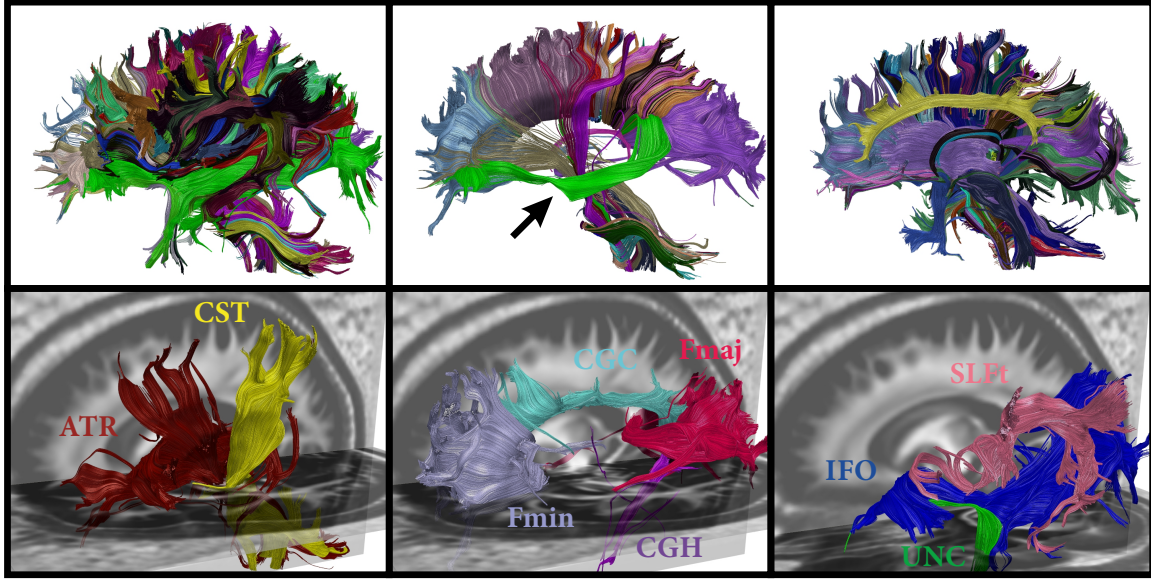


Figure 6.3: Atlas-guided clustering results for one dataset, clustered with the *combined distance measure*. In the top row all bundles are presented. Fiber bundles are shown for the left hemisphere (left image), bundles that cross the hemispheres (middle image) and bundles of the right hemisphere (right image). Different fiber bundles are displayed in distinct colors. In the middle image a tracking error is present that resulted in a fiber bundle connecting the prefrontal lobe and the corpus callosum (green bundle, marked with an arrow). The bottom row shows the clusters that correspond to the atlas classes shown in Figure 6.2.

colored identically. The matched bundles that correspond to the atlas classes in Figure 6.2 are displayed in the bottom row of Figure 6.3 with the same coloring as in Figure 6.2.

To assess the quality of the clustering quantitatively, the spatial agreement (see Eq. 3.12) between matched bundles and atlas classes was determined for all 46 clustered datasets. The spatial agreement for CD, HD and MPD is shown in Figures 6.4–6.6. The figures use box plots to present the spatial agreement of the individual, clustered bundles, obtained with atlas-guided CATSER (in red), CATSER (in green) and HAC (in blue). The centerline in the boxes denotes the median (second quartile). The bottom and the top of a box correspond to the first and the third quartile and define the *inter-quartile range* (IRQ). The lines that emerge from the top and bottom of a box are whiskers and contain all remaining data points that are in the range of $\pm 1.5 \times \text{IRQ}$. All data points that are outside of the whiskers are outliers and marked with grey dots.

For all similarity measures (CD, HD, MPD) the clustering techniques were able to group fibers into bundles that had a high spatial agreement with the classes in the atlas (see Figures 6.4–6.6). Nonetheless, various differences between the three clustering methods were observed for all similarity measures. For every clustered bundle, the variability of the spatial agreement for atlas-guided CATSER clustering is considerably lower compared to clustering with CATSER and HAC. Especially the IRQs are significantly reduced, but also is the number of outliers. By comparing the spatial agreement between CATSER (without atlas) and HAC no striking differences of the IRQs and outliers are apparent.

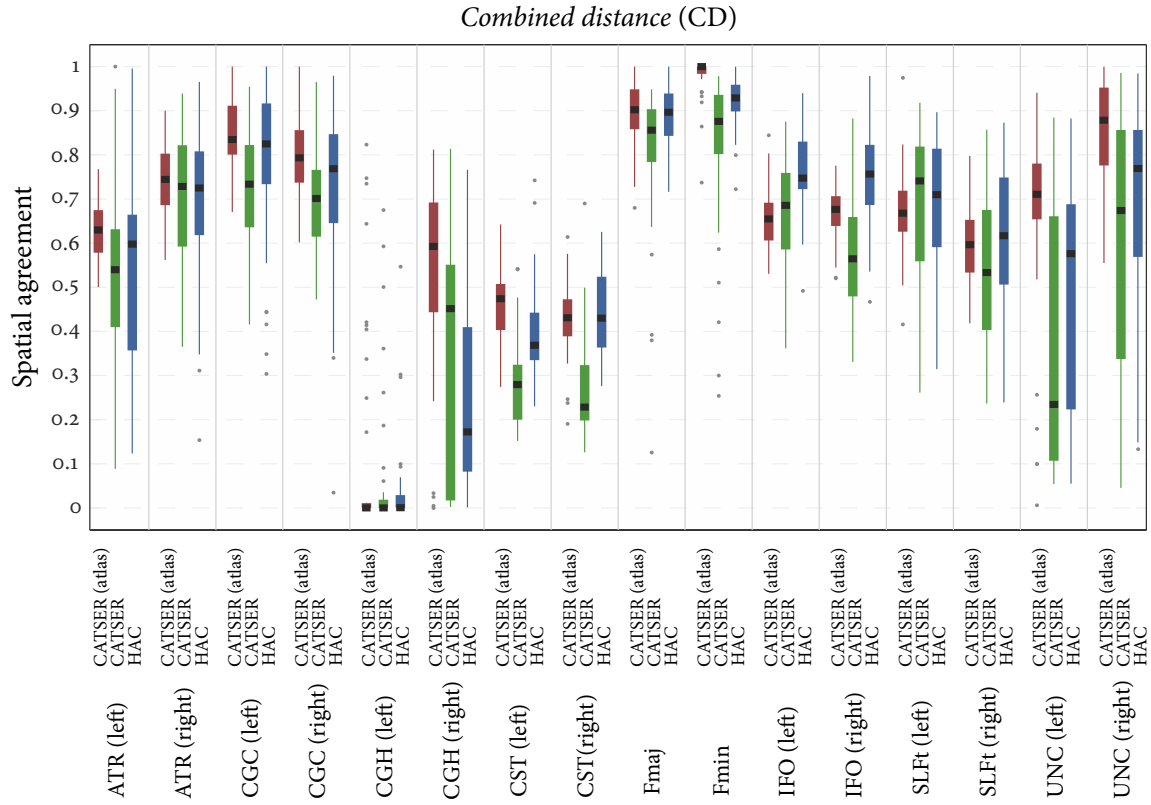


Figure 6.4: Spatial agreement of clustered fiber bundles and atlas classes for the *combined distance* (CD). Using the atlas-matched fiber bundles of the 46 datasets that were clustered with the three different methods (atlas-guided CATSER, CATSER, HAC), the average spatial agreement between fiber bundles and atlas class was determined. The results for each bundle and each clustering technique are shown above.

By inspecting the average spatial agreement of atlas-guided CATSER and CATSER (without atlas) using CD and HD, a higher spatial agreement can be observed for the atlas-guided clustering in most bundles (except for the left IFO (for CD and HD), the left SLFt (for CD) and the right ATR (for HD)). In contrast, by using MPD as similarity measure results are more diverse and spatial agreement was higher for only eight bundles, using atlas-guided CATSER clustering. For the conventional clustering with HAC, higher spatial agreement was observed in certain bundles than in corresponding bundles that were clustered with atlas-guided CATSER. However, it has to be considered that HAC clustered only a small subset of 10 000 tracts, which resulted in considerably smaller clusters that occupied less space compared to the two other techniques. As fewer tracts exist to delineate the whole bundle and to depict its smaller details, a higher spatial correspondence was obtained for HAC. This effect can be easily identified in Figure 6.7, which shows an example of a single fiber bundle (inferior fronto-occipital fasciculus – IFO), clustered with all methods. Using HAC, the best matching bundles contain only few fiber tracts. These few tracts travel primarily through the atlas class, which results in a high spatial agreement. Regarding the other methods, the fiber bundles are more detailed and contain more fiber tracts. They resemble the atlas class

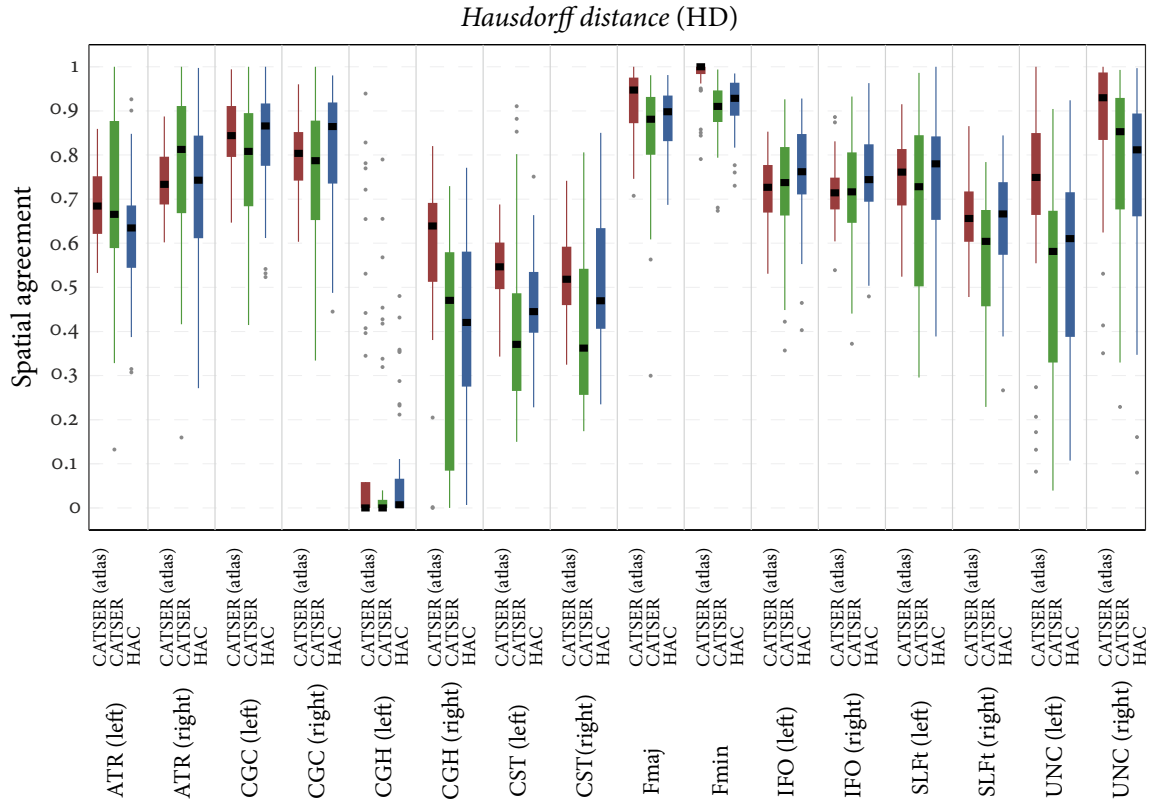


Figure 6.5: Spatial agreement of clustered fiber bundles and atlas classes for the *Hausdorff distance* (HD). Using the atlas-matched fiber bundles of the 46 datasets that were clustered with the three different methods (atlas-guided CATSER, CATSER, HAC), the average spatial agreement between fiber bundles and atlas class was determined. The results for each bundle and each clustering technique are shown above.

more closer, even though the spatial agreement is lower compared to HAC (except for the atlas-guided CATSER clustering with HD). For atlas-guided CATSER with CD, however, the obtained cluster for the IFO is rather large and contains large parts of the uncinate fascicle compared to the clusters obtained with the other methods. By using CATSER (without atlas) and the MPD, the obtained IFO bundle was only partially covered by tracts. While the results for different clustering methods were consistent for all three similarity measures, the spatial agreement with HD was consistently higher for most bundles compared to CD and MPD (Figure 6.4 vs. Figure 6.5 vs. Figure 6.6). Overall, the previous observations are confirmed by Table 6.3, where the average spatial agreement and the standard deviation are summarized for all clustering techniques and similarity measures. For atlas-guided CATSER, HD achieved the highest spatial agreement, followed by CD and MPD. By performing the clustering with CATSER (without atlas), HD still achieved the highest spatial agreement, but is followed by MPD and CD. For the clustering with HAC, the highest spatial agreement was obtained with MPD, followed by HD and CD.

For the CGH (in particular the CGH of the left hemisphere) fiber bundles could not be extracted for all datasets. By analyzing the data, it was observed that no or only few fibers

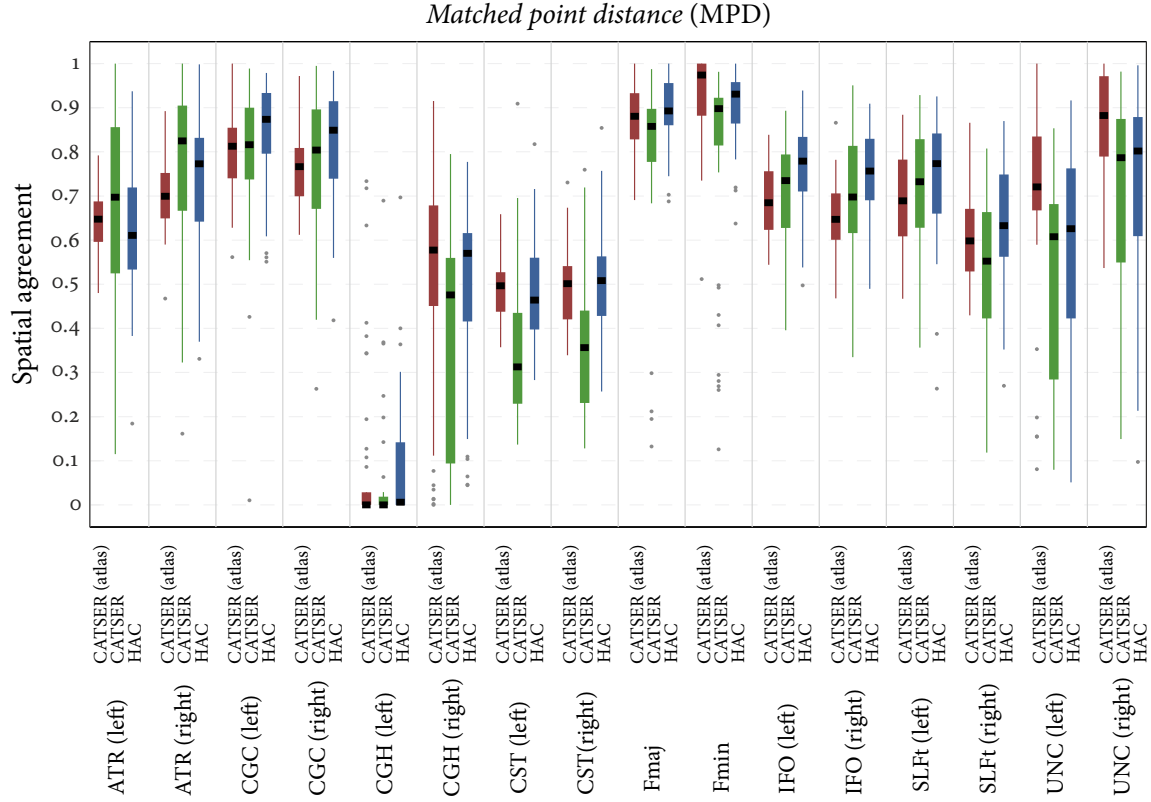


Figure 6.6: Spatial agreement of clustered fiber bundles and atlas classes for the *matched point distance* (MPD). Using the atlas-matched fiber bundles of the 46 datasets that were clustered with the three different methods (atlas-guided CATSER, CATSER, HAC), the average spatial agreement between fiber bundles and atlas class was determined. The results for each bundle and each clustering technique are shown above.

traversed the volume of the CGH atlas classes for certain datasets. This is also captured in Figures 6.4 and 6.5, in which the CGH bundle has a very low spatial agreement and a high spread of values. This is presumably due to imperfections in the coregistration, the small volume of the CGH atlas class and the small number of tracts that this bundle contains. Some of the very few tracts that belonged to the CGH were presumably removed by the outlier elimination. The corticospinal tract also has a relatively low spatial agreement (Figures 6.4 and 6.5), which is the result of merging the CST bundle with adjacent tracts that are not part of the CST but are highly similar. While the CST is a very narrow fiber bundle, the tracts in its vicinity are still highly similar compared to the tracts that constitute the CST. Even though anatomical information was incorporated into the cluster analysis, some highly similar tracts are still merged into the CST, which leads to a reduction in spatial agreement (compare the occupied volume of the CST atlas class and the corresponding CST cluster in Figures 6.2 and 6.3).

With the atlas-guided CATSER clustering, an elevated IRQ of the spatial agreement was observed for certain bundles (e.g., UNC). This increase is primarily associated with the inclusion of additional tracts during clustering that belong only partly to the bundle. In Figure 6.8,

one example is presented that shows the uncinate fasciculus (UNC) of the left hemisphere for two datasets and the volume of the corresponding atlas class (in semi-transparent green). While the tracts follow the anticipated trajectory of the fibers in the clustered UNC in the left image, the bundle in the right image contains additional fibers that leave the bundle (arrows). These incorrect tracts share major parts of the atlas-class volume and also have a high partial correspondence to other tracts in the bundle.

In order to demonstrate the consistency of the atlas-guided CATSER clustering, Figure 6.9 displays the temporal part of the superior longitudinal fasciculus (SLFt) for different datasets. The clusters are superimposed on the volume of the corresponding atlas class. Even though clusters vary between subjects, the appearance, the shape and size of the clusters are very similar and close to the volume-based atlas representation. Due to the fact that various sub-parts of the SLFt leave the atlas class, the spatial agreement is decreased and ranges between 0.60 and 0.82 for the shown bundles.

During the analysis of the clustered datasets, some tracking errors were observed. In Figure 6.3, for example, various tracts were traced by the tracking algorithm that connect the prefrontal lobe and the corpus callosum (green bundle in the middle image of the top row, marked with an arrow). Another tracking error is visible in Figure 6.9, in which the SLFt bundle of various datasets connect to the external capsule in the sub-insular white matter (see regions of the bundles marked with an arrow).

6.5.2 Outlier elimination

Results of the outlier elimination are presented in Figure 6.10. The fiber tracking dataset in the upper left corner was manually segmented and the 16 fiber bundles that are also defined in the atlas were extracted. The data was not spatially transformed and resided in its native space. For visualization of the data, a tableau view shows the data in three different orientations: top left image = anterior-posterior view, bottom left image = left-right view, top right image = superior-inferior view. Each of the segmented fiber bundles is shown in a different color. For all tracts of the dataset the centroids were computed and used as gold standard. Centroids are shown in the top right corner, with the same tableau view and identical coloring as the tracts.

To investigate the clustering with CATSER, the centroids were clustered with different outlier elimination strategies (low, moderate, high outlier elimination) and varying levels of

Table 6.3: Average spatial agreement and standard deviation for all clustering techniques and similarity measures.

Similarity measure	Clustering technique		
	atlas-guided CATSER	CATSER	HAC
CD	0.66 ± 0.21	0.54 ± 0.22	0.60 ± 0.23
HD	0.70 ± 0.20	0.63 ± 0.22	0.65 ± 0.21
MPD	0.65 ± 0.20	0.60 ± 0.22	0.66 ± 0.20

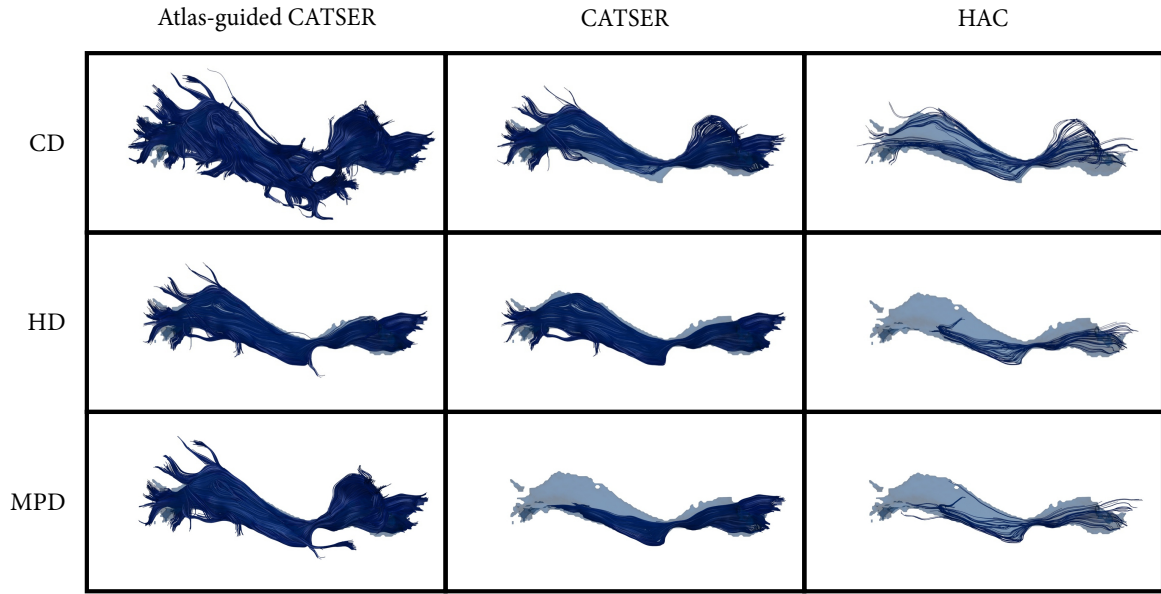


Figure 6.7: Inferior fronto-occipital fasciculus (IFO) of one dataset clustered with all three methods (atlas-guided CATSER, CATSER, HAC) and all similarity measures (CD, HD, MPD). Bundles are shown in the atlas space and are superimposed on the corresponding class of the atlas (in semi-transparent blue). The spatial agreement for bundles clustered with CD is (from left to right): 0.58; 0.62; 0.69, for bundles clustered with HD: 0.75; 0.67; 0.73 and for bundles clustered with MPD: 0.65; 0.69; 0.69.

noise ($D_1=0\%$, $D_2=33\%$, $D_3=66\%$, $D_4=99\%$). The results for this tract-centroid clustering are shown below the gold standard datasets (tracts, centroid). Again, the tableau view is used to visualize the results of the clustering in three distinct orientations. Resulting clusters are shown in different colors, while the colors of the clusters were selected in correspondence to the coloring of the gold standard dataset (top right tableau in Figure 6.10). Centroids that were identified as outlier during the clustering are displayed in grey.

First of all, the transformation from the segmented tracts (top left) to centroids (top right) already resulted in some isolated centroids that are visible in the tableau of the (segmented) fiber tract centroids (the gold standard).

By inspecting the clustered tableaus in Figure 6.10, it is evident that the quality of the clustering depends on both the level of noise added to the tract centroids (see section 6.4.2) and the outlier elimination strategy. Overall, the framework was able to handle all scenarios quite well. However, no outlier elimination strategy was able to achieve an adequate clustering in all four noise scenarios ($D_1 \dots D_4$). If noise is barely present (D_1 , D_2), minimal outlier elimination is sufficient to obtain a good discrimination between the clusters, while a more aggressive outlier elimination strategy is indispensable if the data contains too much noise (D_3 , D_4). If more aggressive outlier elimination strategies are used for datasets with little noise (e.g., D_1 , D_2), the outlier elimination rate might be too high and small clusters might be labeled as outliers. In the opposite case, when plenty of noise is present (e.g., D_3 , D_4) and minimal outlier elimination is used, clustering will probably fail and lead to erroneous clusters.

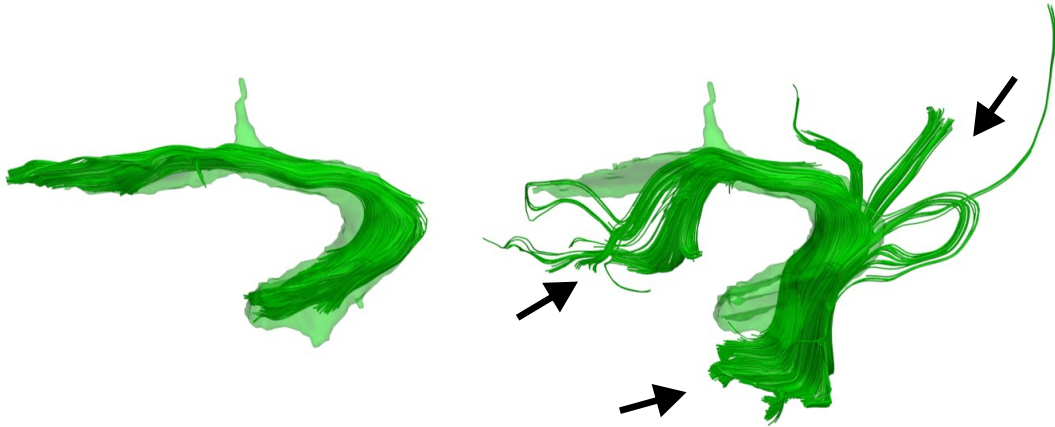


Figure 6.8: Left uncinatus fasciculus (UNC) of two datasets clustered with atlas-guided CATSER clustering. Both bundles reside in the atlas space and are superimposed on the corresponding class of the atlas (in semi-transparent green). While the bundle on the left follows the anticipated trajectory of the UNC, the bundle on the right side contains additional tracts that leave the bundle and follow other paths. The spatial agreement for the left bundle is $s = 0.93$ and for the right bundle $s = 0.58$.

In general, the number of points that were identified as outliers increases with intensified outlier elimination. Hence, it seems that a moderate outlier elimination strategy is a good compromise.

Nevertheless, an increase in noise always has harmful effects on clustering and will result in degraded cluster quality. The clustering methods will have difficulties to identify the correct clusters and chaining effects will lead to the merging of distinct clusters. The more noise being present the more aggressive the outlier elimination should be in order to achieve a satisfying clustering with good delineation of fiber bundles.

6.5.3 Performance analysis

Performance analysis of the clustering framework

The performance analysis of the clustering framework was conducted as described in section 6.4.3. Results of the analysis for the individual parts of the clustering framework and the analyzed similarity measures (CD, HD, MPD) are presented in Figure 6.11. The overall computation time in dependency of the number of employed CPU cores is shown with stack plots on the left side, along with the corresponding speedup on the right side.

Overall, the clustering of the data using the MPD similarity measure (Figure 6.11e) had the lowest computation time followed by the clustering with CD (Figure 6.11a) and HD (Figure 6.11c). Due to the separation of the similarity measures from the clustering, precomputation of the similarity measures (part 1, step 2 in Figure 3.2) and assignment of remaining tracts to the prototype clusters (part 3, step 8 in Figure 3.2) were the most time consuming parts that accounted for major portions of the computation time. Using CD and HD, the actual clustering process that is responsible for the generation of prototype clusters from the

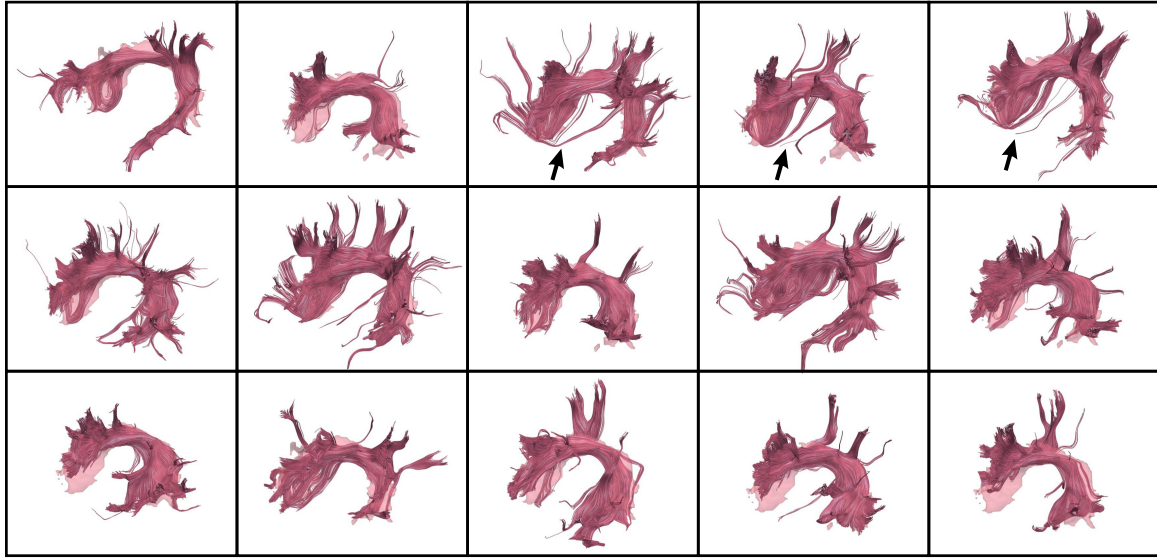


Figure 6.9: Temporal part of the left superior longitudinal fasciculus (SLFt) for 15 datasets. The datasets reside in the atlas space and were clustered with the atlas-guided CATSER clustering. From left to right, the spatial agreement for the bundles in the top row is: 0.62; 0.65; 0.67; 0.61; 0.68, in the middle row: 0.67; 0.60; 0.82; 0.64; 0.60 and in the bottom row: 0.64; 0.61; 0.74; 0.65; 0.69. In certain images, tracking errors are present that connect the SLFt to the external capsule (see arrows).

sample (part 2, steps 3–7 in Figure 3.2) accounted only for a fraction of the total computation time. For MPD, however, the computation time was governed by the CATSER clustering itself (part 2) due to the low algorithmic complexity of MPD.

Increasing the number of active cores led to a significant reduction of the computation time for all similarity measures. The speedup in Figure 6.11b for part 1 and part 2 using CD is fairly linear for an increasing number of CPU cores. For HD, the speedup in Figure 6.11d scales with number of cores, while it scales only moderately for MPD. This implies that the HD could be parallelized very efficiently, whereas the additional CPU cores are not as efficiently used with CD and MPD. For the CATSER clustering (part 2), the increase of partitions that are clustered in parallel resulted in super linear speedup¹ ($S_P > P$) for all three similarity measures (Figures 6.11b, 6.11d and 6.11f).

While the performance analysis was only performed for unsupervised clustering, an increase in computation time of approximately 10 % was observed for atlas-guided CATSER clustering.

Performance analysis for the computation of the *Hausdorff distance*

The performance analysis of the *Hausdorff distance* was conducted with three different GPU systems and one CPU system (see section 6.4.3). Results for the computation time and the speedup using the 32 core CPU system and algorithms 1 and 2 are shown in Figure 6.12. As

¹As a matter of fact, the super linearity of the speedup is the result of two contributing factors: the increase in the number of processing cores and the increase of parallelly clustered partitions.

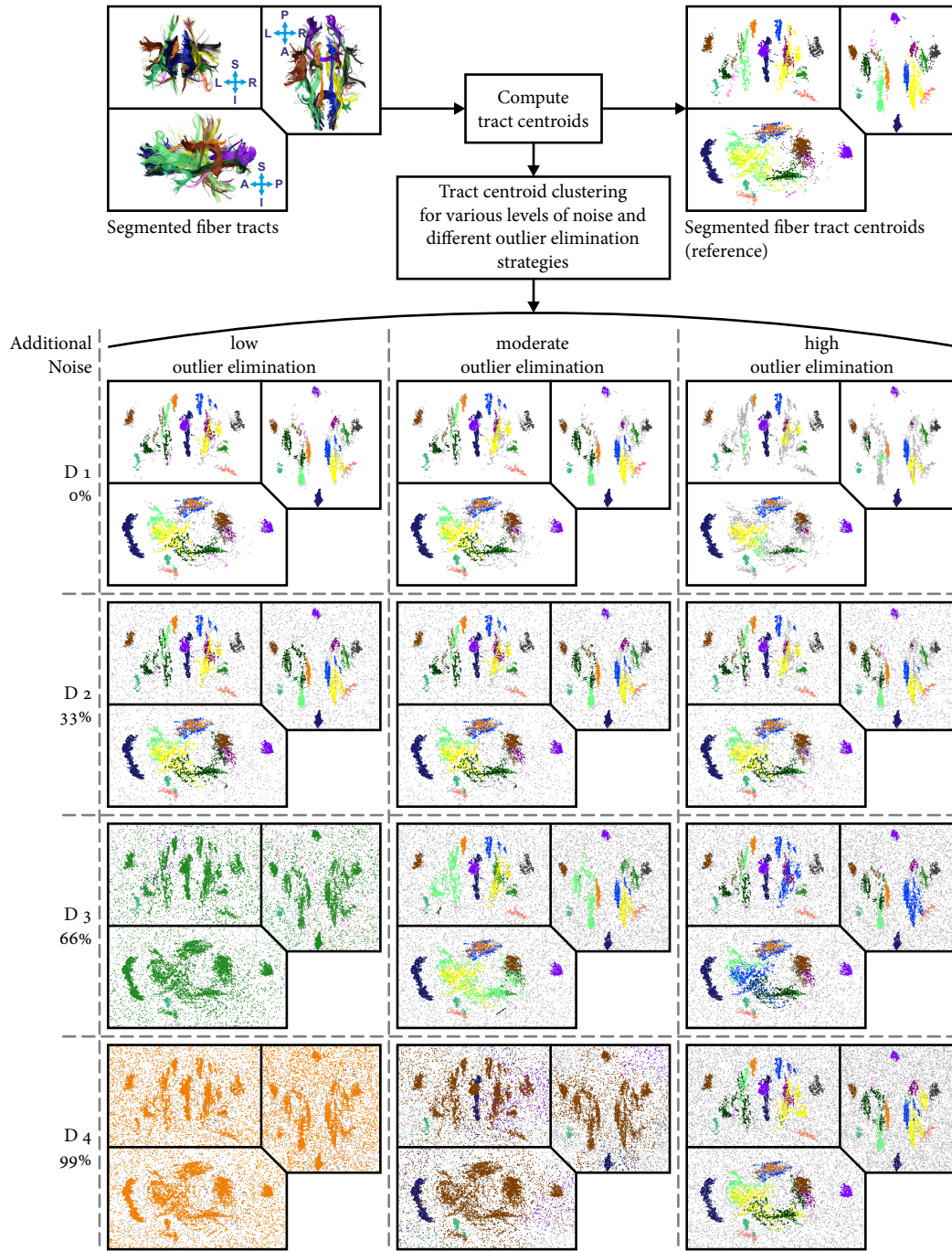
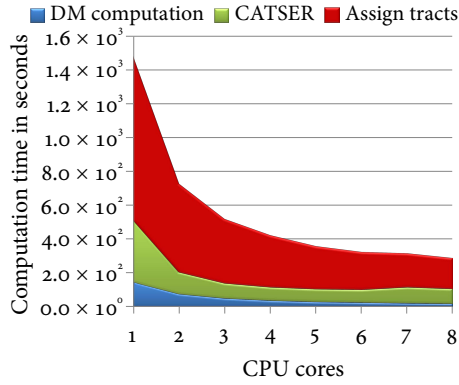
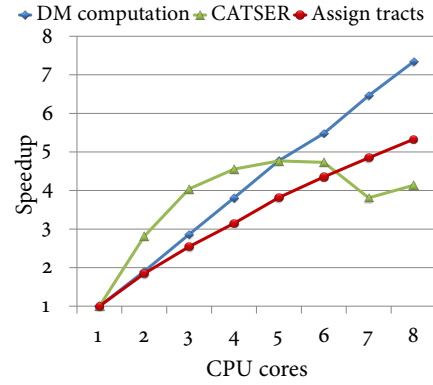


Figure 6.10: Results of the experiments that were conducted to demonstrate the effects of outlier elimination. One dataset was segmented into a set of bundles that are shown in distinct colors in the top left corner. Using the segmented fiber tracts, centroids of tracts were computed and used as a gold standard. Each point in the gold standard image (top right) represents the centroid of one fiber tract. The groups of centroids (gold standard) are shown in the same color as their corresponding fiber bundles. All data is presented in a tableau view that shows the data in three different orientations: top left image = anterior-posterior view, bottom left image = left-right view, top right image = superior-inferior view. The centroids of the gold standard dataset were clustered using the *tract centroid distance* as similarity measure. Varying levels of artificial noise (0 %, 33 %, 66 % and 99 % additional noise) were used and different outlier elimination strategies were tested. The results for the clustering are displayed in the 4×3 table (D1–D4).

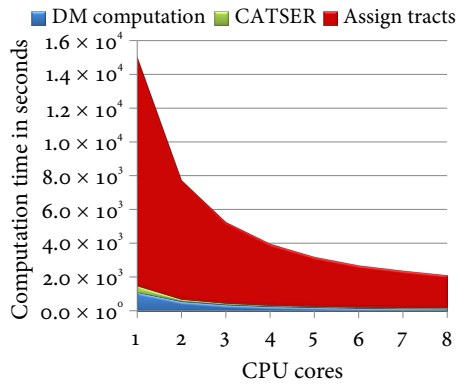
Chapter 6 Clustering experiments



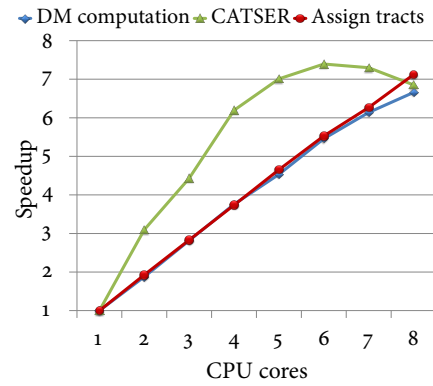
(a) Computation time for CD.



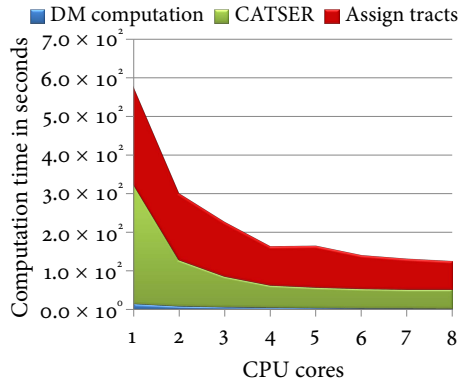
(b) Speedup for CD.



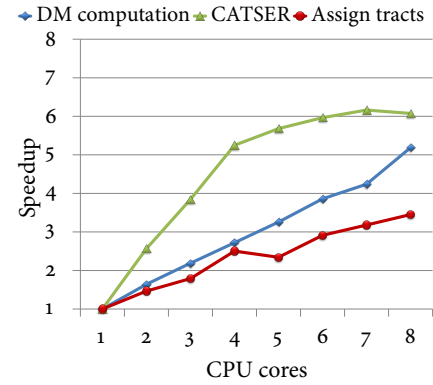
(c) Computation time for HD.



(d) Speedup for HD.



(e) Computation time for MPD.



(f) Speedup for MPD.

Figure 6.11: Performance analysis of the cluster analysis for the similarity measures CD (top row), HD (middle row) and MPD (bottom row). The overall computation time is shown on the left side and the achieved speedup on the right side. For the analysis, the clustering process was separated into three distinct parts and analyzed separately (see section 6.4.3). In all plots, each part is highlighted in distinct colors, whereas blue denotes the computation of the distance matrix (part 1: "DM computation"), green the clustering (part 2: "CATSER") and red the labeling of remaining tracts (part 3: "Assign tracts"). For CD and HD, part 1 and part 3 were the most time consuming stages. By utilizing multiple CPUs a high speedup was achieved and computation time was reduced significantly. While the speedup is nearly optimal for HD, it is reduced for CD and MPD. For MPD, the computation time is already low due to the reduced algorithmic complexity of the MPD. Compared to the CD and HD, it was therefore not governed by the computation of the similarities (parts 1 and 3) but by the clustering itself (part 2).

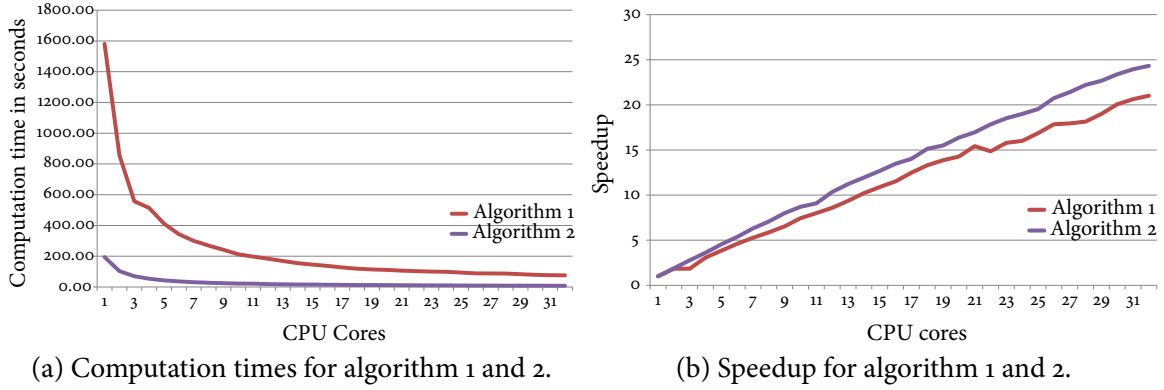


Figure 6.12: Computation times and speedup for the *Hausdorff distance* (algorithm 1 and 2), computed with the 32 core CPU system.

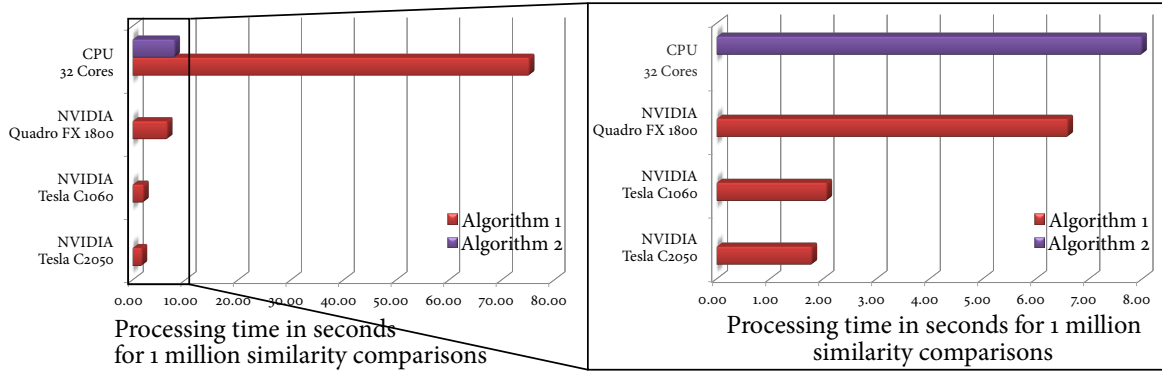


Figure 6.13: Computation times for the *Hausdorff distance* with different GPU systems and the 32 core CPU system.

already observed in the previous section, the clustering framework achieves good speedups for both algorithms with the CPU systems, even if many processing cores are employed (cmp. Figure 6.12). However, severe differences were observed in the computation times between the two algorithms. In comparison to algorithm 1, algorithm 2 had a dramatically lower processing time and a slightly higher speedup.

The processing times for 1 million similarity computations with the GPU systems and the CPU system (using all 32 cores, CPU system 2) are presented in Figure 6.13. Even though algorithm 1 was solely implemented for GPUs, the processing times with all GPU systems were significantly reduced compared to the CPU system. Algorithm 1 on the GPU was not only faster than algorithm 1 on the CPU but also faster than algorithm 2 on the CPU. Even the weakest GPU system (Quadro FX 1800, GPU system 1) outperformed the 32 core CPU system. Despite the fact that the GPU system C2050 has almost twice the number of CPU cores compared to the C1060, differences between both systems were quite small. By comparing computation times between the fastest GPU (C2050, GPU system 3) and the fastest CPU (32 cores, CPU system 2), a speed up of ≈ 40 was observed for algorithm 1 and a speed up of ≈ 3.5 for algorithm 2.

Part III

Quantitative analysis

7

Chapter 7

Quantitative analysis of fiber bundles

What we observe is not nature itself, but nature exposed to our method of questioning.

— Werner Heisenberg (1901–1976)

Quantitative *fiber bundle-driven techniques* are promising methods to investigate altered diffusion properties in individual fiber bundles for group-based imaging studies. In this chapter, a new approach is presented that uses fiber bundles to enhance quantitative analyses. Previously presented techniques are briefly discussed before concepts to improve the analysis are introduced. To investigate the peculiarities of the proposed method, various experiments were performed.

7.1 Introduction

Alterations in brain diffusivity are often indications for changes in the structural organization of the brain. In order to study subtle changes of the white matter microstructure between groups of subjects, a quantitative statistical analysis is frequently performed. While established methods such as *voxel-based morphometry* – VBM (Ashburner and Friston, 2000) or *tract-based spatial statistics* – TBSS (Smith et al., 2006) are predominately used to investigate diffusion in the whole brain or the white matter skeleton, *fiber bundle-driven techniques* (FDTs) facilitate quantitative analysis of diffusion properties in individual fiber bundles (see section 2.5). The earliest concepts to study characteristics of selected white matter pathways with fiber tracts were presented by Fillard et al. (2003) and later extended by various researchers (e.g., Gong et al. (2005); O'Donnell et al. (2007); Hua et al. (2008); Yeatman et al. (2012)). All of these tractography-based approaches investigate diffusion properties along the course of specific *bundles of interests* (BOIs). Often, a mean tract projection is employed in which quantitative values along the tracts of the BOI (e.g., FA, ADC) are projected onto corresponding points of the BOI's mean tract¹. Even though this reduces data dimensionality

¹A mean tract is either a constructed tract that follows the average course of the bundle or a tract of the bundle with the highest average similarity to all other tracts.

from three dimensions to one dimension and results in a more convenient analysis of diffusion data, these approaches have various limitations.

Due to the analysis of data along the course of bundles and the reduction of data dimensionality, these methods are restricted to analyze BOIs with a simple tubular structure. If BOIs are not tubular but fan out, split up or even contain misguided tracts with incorrect trajectories, the statistical analysis of diffusion properties along the course of the bundles will be compromised. If the mean tract approach is employed, significant differences in distinct regions of the bundle perpendicular to the mean tract may remain undetected. In addition, bundle-specific analysis with FDTs is obviously limited to bundles that can be traced reliably and separated correctly from other bundles (Smith et al., 2006).

Another important but often neglected aspect are adverse, normalization-induced interpolation effects at the boundaries of grey and white matter structures. As demonstrated by Chao et al. (2009), interpolation not only affects grey and white matter boundaries but also boundaries between distinct white matter structures. If contributions of different bundles are not separated, statistically incorrect results may arise. As these effects are only present if spatial normalization is used to align and superimpose different datasets, a solution to avoid normalization-induced interpolation is to perform the quantitative analysis in native space. With this latter approach, however, spatial correspondence between fiber bundles of different datasets has to be established with alternative techniques. To perform valid statistical analyses with FDTs, these problems have to be taken into account, which often requires manual, time-consuming interventions of an operator.

In this chapter, a new method to improve the quantitative *fiber bundle-driven analysis* of cerebral diffusion weighted MRI data is presented. Instead of studying diffusion properties solely along the course of specific white matter bundles, a voxel-based approach is employed that analyzes diffusion properties in entire fiber bundles. To this end, the clustering framework (see Part II) is used to automatically partition fiber tracts into bundles that correspond to known anatomic white matter structures. To avoid adverse interpolation effects at the boundaries of tissue structures, quantitative values are projected onto the tracts before spatial normalization. The analysis is then performed independently for each bundle while suppressing contributions of voxels that do not belong to the selected white matter structures.

7.2 Quantitative analysis of fiber bundles

7.2.1 General workflow

In the following sections, a new FDT for analyzing diffusion MRI data is outlined and the individual processing steps are described. Compared to other FDTs, the analysis is not performed along the trajectory of the fiber bundle but uses the voxels of the entire bundle. The method is therefore not restricted to tubular structures and can be applied to study complex fiber pathways (e.g., bundles that fan out or split up).

The presented technique also avoids normalization-induced interpolation between distinct anatomical structures that may occur as a result of the spatial normalization of different datasets (Chao et al., 2009). In Figure 7.1, normalization-induced interpolation is illustrated

with two different white matter structures. Applying spatial normalization may result in an overlap of bundles and interpolation of quantitative values in overlapping areas.

To avoid this and additionally suppress contributions of other anatomical structures, fiber tracts that follow the course of white matter structures are instrumented. As fiber tracts are defined by real-valued points in 3D space (see section 2.4), the spatial normalization of these points results in no interpolation but a coordinate transformation that changes the course of the tracts. By utilizing this tract-specific property, interpolation can be avoided if quantitative values are projected onto the tract points prior to the normalization. As the spatial transformation changes only the spatial location of tract points, quantitative values that are attached to these points are preserved. By organizing fiber tracts into bundles that correspond to anatomically known white matter structures and rasterizing each bundle separately, corruption of quantitative values by other bundles is prevented. This concept is illustrated in Figure 7.2 by using the previously employed white matter structures (see Figure 7.1). Following separate rasterization and gridding of bundles, a voxel-wise quantitative analysis is performed individually for each fiber bundle. Due to the individual extraction of fiber bundles for each dataset, reliability maps can be computed for each fiber bundle which allows exclusion of regions to which only few datasets contribute.

The entire processing workflow for the quantitative analysis of multiple datasets is illustrated in Figure 7.3. The seven workflow steps are briefly divided into four major parts that are outlined in the following sections. The projection of quantitative values onto the tracts (step 1) is described in section 7.2.2 before section 7.2.3 covers spatial normalization and grouping of tracts (steps 2–4). The gridding of attached values (step 5) is outlined in section 7.2.4 and the final voxel-wise statistical analysis of individual fiber bundles (steps 6 and 7) in section 7.2.5.

Conventional processing

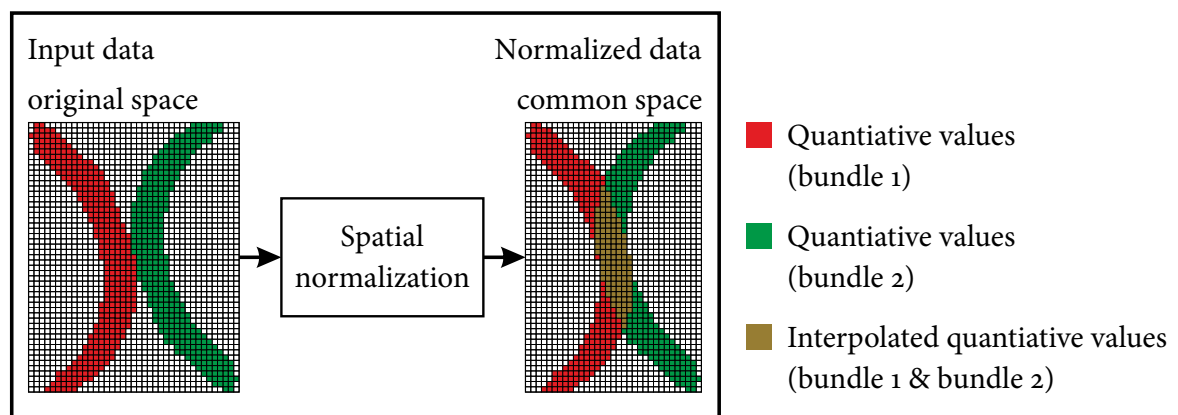


Figure 7.1: Illustration of the normalization-induced interpolation of conventional methods for quantitative analysis. The red and green regions correspond to quantitative values of two distinct white matter structures on a regular grid. Due to spatial normalization of the individual datasets and the resulting dislocation of the structures, interpolation between white matter bundles may occur.

Fiber bundle-based processing

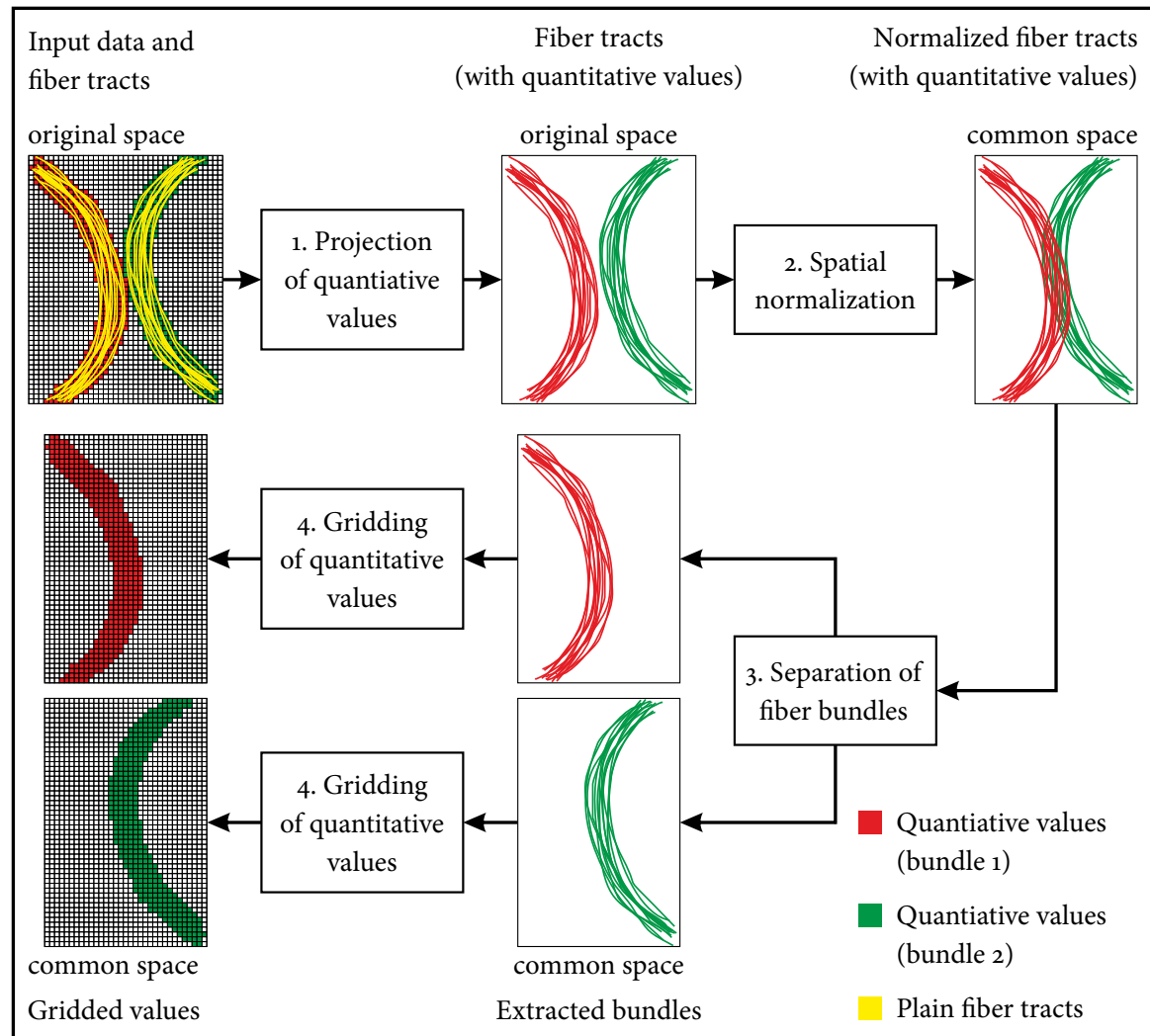


Figure 7.2: Illustration of fiber bundle-based processing that avoids normalization-induced interpolation. Again, red and green regions correspond to quantitative values of two distinct white matter structures. Fiber tracts that traverse through these regions are shown as lines. Yellow lines denote tracts without quantitative values, while red and green lines correspond to tracts with attached, spatially corresponding quantitative values. In order to avoid interpolation of quantitative values, the white matter structures are analyzed independently. For this purpose, quantitative values are projected onto the fiber tracts prior to the spatial normalization (step 1, tracts in red and green). Spatial normalization results in a coordinate transformation (step 2), which preserves the quantitative values that are attached to the tracts. After spatial normalization, fiber tracts are partitioned into bundles that correspond to anatomical structures (step 3). By processing fiber bundles independently, contributions of different bundles are masked out. For the quantitative analysis, values that are still attached to the fiber tracts are finally gridded to a regular grid in common space (step 4).

7.2 Quantitative analysis of fiber bundles

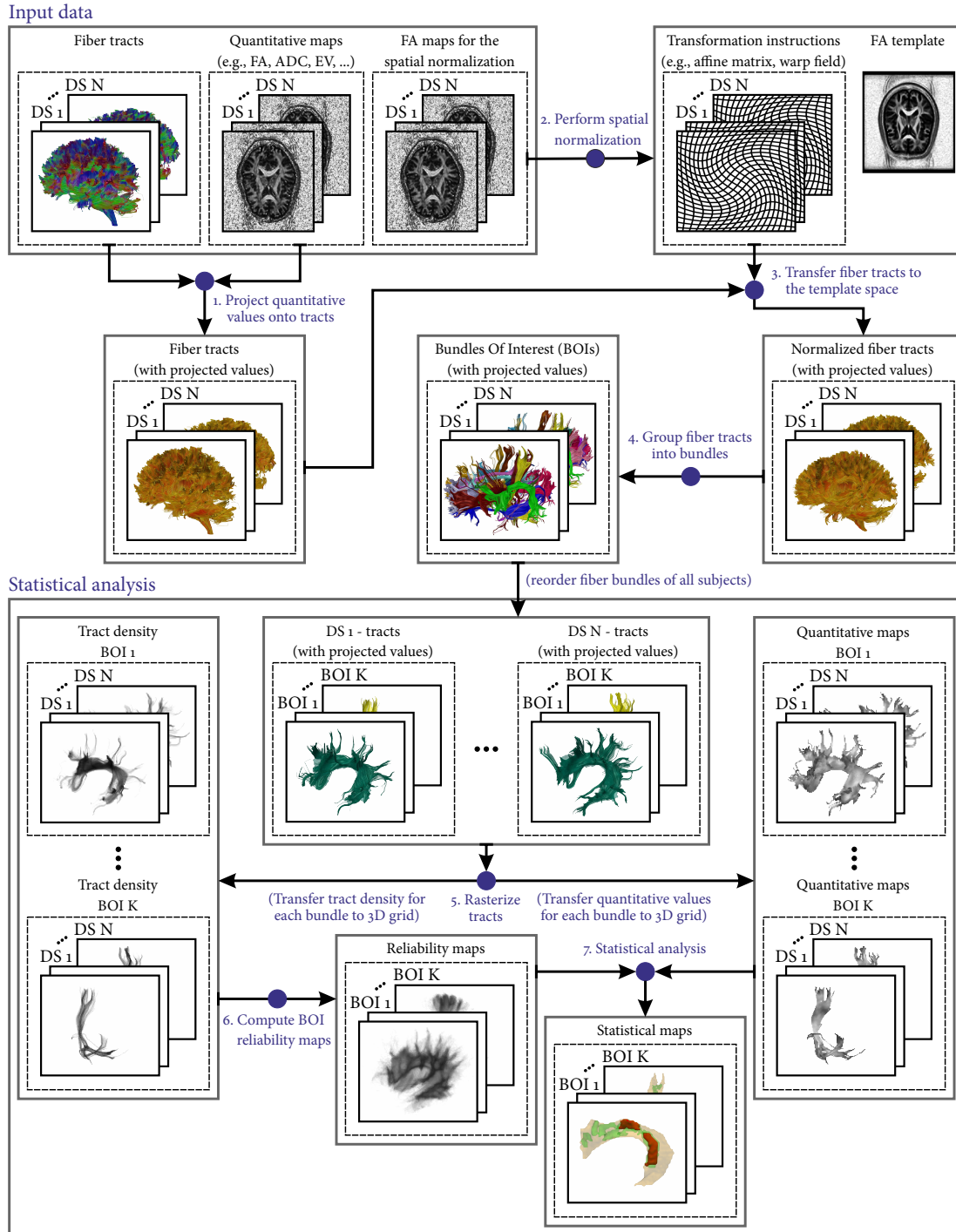


Figure 7.3: Workflow of the quantitative fiber bundle-driven analysis. After data acquisition and preprocessing, individual quantitative values are projected onto the tracts of each dataset (step 1). Using the FA maps of all datasets, a spatial normalization is performed (step 2) and fiber tracts are transferred to the normalized space (step 3). For each dataset, fiber tracts are grouped into bundles (step 4). Tracts are rasterized to obtain tract density and quantitative maps for all fiber bundles (step 5). Tract density maps are then used to create reliability maps that determine how reliably a voxel can be associated with the fiber bundle (step 6). Using reliability and quantitative maps, statistical analyses are performed to detect significant differences in each BOI (step 7).

7.2.2 Projection of quantitative values

Projection of quantitative values onto fiber tracts prior to the spatial normalization is an important step to avoid normalization-induced interpolation.

For all tract points that define the trajectories of the tracts in 3D space, quantitative values are determined and attached to these tract points. To compute the quantitative value $\tau(\mathbf{p}_j)$ that is projected onto an arbitrary tract point \mathbf{p}_j , trilinear interpolation is used which takes into account the quantitative values of all voxels $\{\mathbf{v}_1, \dots, \mathbf{v}_8\}$ that surround \mathbf{p}_j (see Figure 7.4). If $\tau(\mathbf{v}_i)$ denotes the quantitative value of voxel $\mathbf{v}_i \in \{\mathbf{v}_1, \dots, \mathbf{v}_8\}$, the quantitative value $\tau(\mathbf{p}_j)$ for \mathbf{p}_j is given by:

$$\tau(\mathbf{p}_j) = \sum_{i=1}^8 \tau(\mathbf{v}_i) \times k_i(\mathbf{v}_i, \mathbf{p}_j), \quad (7.1)$$

whilst $k(\mathbf{v}_i, \mathbf{p}_j)$ denotes the weighted contribution of voxel \mathbf{v}_i .

In order to determine the weighting value $k_i(\mathbf{v}_i, \mathbf{p}_j)$ for a tract point \mathbf{p}_j and a voxel \mathbf{v}_i , voxels are interpreted as point sources that are located at the vertices of a unit cube (Figure 7.4). From this, it follows that the x, y and z coordinates of all voxels $\{\mathbf{v}_1, \dots, \mathbf{v}_8\}$ along the three axes are either 0 or 1. Hence, the voxels have the following coordinates:

$$\begin{aligned} \mathbf{v}_1 &= (0, 0, 0), \mathbf{v}_2 = (0, 0, 1), \mathbf{v}_3 = (0, 1, 0), \\ \mathbf{v}_4 &= (0, 1, 1), \mathbf{v}_5 = (1, 0, 0), \mathbf{v}_6 = (1, 0, 1), \\ \mathbf{v}_7 &= (1, 1, 0), \mathbf{v}_8 = (1, 1, 1). \end{aligned} \quad (7.2)$$

It is further assumed that tract point $\mathbf{p}_j = (p_j^x, p_j^y, p_j^z)$ is located inside the cube and its coordinates p_j^x, p_j^y, p_j^z are bounded by $[0, 1]$ with $0 \leq p_j^x \leq 1$, $0 \leq p_j^y \leq 1$ and $0 \leq p_j^z \leq 1$. Then, $k_i(\mathbf{v}_i, \mathbf{p}_j)$ is the product of \mathbf{p}_j 's projections onto each axis x, y and z . If \mathbf{v}_i is a voxel with coordinates v_i^x, v_i^y, v_i^z , the value $k_i(\mathbf{v}_i, \mathbf{p}_j)$ is given by:

$$k_i(\mathbf{v}_i, \mathbf{p}_j) = \delta_x(v_i^x, p_j^x) \times \delta_y(v_i^y, p_j^y) \times \delta_z(v_i^z, p_j^z), \quad (7.3)$$

where $\delta_x(v_i^x, p_j^x)$ denotes the contribution of \mathbf{p}_j along the x axis in dependency of v_i^x :

$$\delta_x(v_i^x, p_j^x) = \begin{cases} p_j^x & \text{if } v_i^x = 1, \\ 1 - p_j^x & \text{if } v_i^x = 0. \end{cases} \quad (7.4)$$

For the remaining axes, the contributions $\delta_y(v_i^y, p_j^y)$ and $\delta_z(v_i^z, p_j^z)$ are computed analogously.

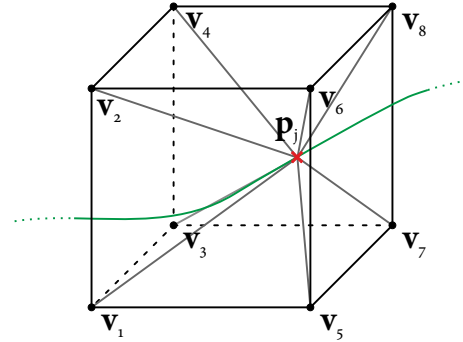


Figure 7.4: Quantitative values are attached to the fiber tracts with trilinear interpolation. For each tract point \mathbf{p}_j , the nearest neighboring voxels $\{\mathbf{v}_1, \dots, \mathbf{v}_8\}$ are determined and their quantitative values are used to approximate the quantitative value at the intermediate location of \mathbf{p}_j . Hereby, surrounding voxels are interpreted as point sources that are located at the vertices of a unit cube.

7.2.3 Spatial normalization and separation of fiber bundles

In order to perform the quantitative analysis, a spatial normalization is performed that aligns and superimposes corresponding anatomical structures of different datasets. During this normalization process, the different datasets are distorted and transferred to a common template space (see e.g., section 6.1). As the normalization of the tracts results in a coordinate transformation of the tract points, values that are attached to the points are preserved.

To suppress contributions of other structures during the statistical analysis, the white matter structures are analyzed individually by using fiber tracts. As fiber tracts have no apparent structural organization and are loosely distributed throughout the brain, it is uncertain to which underlying white matter structure particular tracts belong to. It is therefore necessary to group tracts into fiber bundles that represent the underlying white matter anatomy².

To delineate the bundles, manual methods are often used that are highly time consuming, prone to errors and require an operator with profound anatomical knowledge (see section 2.5). However, as datasets are spatially normalized and transferred to common template space, automatic techniques such as the previously presented atlas-based clustering are highly suitable to group tracts automatically into fiber bundles (see chapter 3).

7.2.4 Gridding of quantitative values

To perform the voxel-wise analysis, tracts are rasterized and quantitative values that are still attached to the tracts are transferred to a regular grid with a user-defined target resolution.

To separate distributions from distinct bundles and avoid the interpolation of quantitative values that belong to different white matter structures, gridding is performed independently for each fiber bundle. All tracts that belong to the bundle are rasterized and quantitative values are gridded to the voxels they occupy. If multiple tracts of a bundle contribute to the quantitative value of a single voxel, contributions of the different tracts are averaged.

With the rasterization of a tract p , quantitative values are gridded to voxels that are traversed by p . If a single fiber tract p traverses through an arbitrary voxel v_i , the quantitative value $\tau(v_i)$ in voxel v_i has to be determined. To compute $\tau(v_i)$, two cases are considered that depend on the number of tract points $\rho(p, v_i)$ that occupy voxel v_i :

$$\tau(v_i) = \begin{cases} \tau(p_j) & \text{if } \rho(p, v_i) = 1, \\ \tau(p_j) \times \frac{\|v_i - p_{j+1}\|}{\|p_j - p_{j+1}\|} + \tau(p_{j+1}) \times \frac{\|v_i - p_j\|}{\|p_j - p_{j+1}\|} & \text{if } \rho(p, v_i) = 0, \end{cases} \quad (7.5)$$

with $p_j, p_{j+1} \in p$. In the first case, a single tract point p_j is located in voxel v_i . Then, the quantitative value $\tau(v_i)$ corresponds to the quantitative value $\tau(p_j)$ that is attached to p_j . In the second case, the number of tract points $\rho(v_i)$ that occupy voxel v_i is zero. Hence, two successive tract points p_j and p_{j+1} are located outside of the voxel and span a line that traverses v_i . Linear interpolation is then used to approximate the quantitative value $\tau(v_i)$ in voxel v_i between p_j and p_{j+1} . If a situation arises in which multiple tract points are located in

²If fiber tracts are not partitioned into bundles and the entire tractography dataset is processed, interpolation of quantitative values will occur later during gridding.

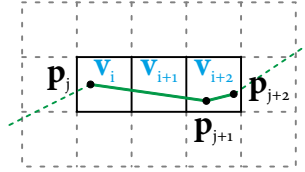


Figure 7.5: After spatial normalization and separation of fiber bundles, quantitative values that are attached to the tract points are gridded to a regular grid in order to separate contributions of distinct white matter structures. For the computation of the quantitative values, different cases are distinguished that depend on the number of tract points that are located inside a single voxel. If exactly one tract point is located inside the voxel, the quantitative value inside the voxel corresponds to the quantitative value of the tract point (e.g., tract point p_j in voxel v_i). If no tract point is located inside the voxel, but two successive tract points span a line that traverses the voxel, linear interpolation is used to approximate the quantitative value at the intermediate voxel position (e.g., tract points p_j , p_{j+1} and voxel v_{i+1}). If a tract contributes multiple times to the quantitative value of a voxel, values are computed separately before they are finally averaged (e.g., tract points p_{j+1} , p_{j+2} and voxel v_{i+2}).

one voxel or a tract travels multiple times through a voxel, quantitative values are computed separately before they are finally averaged. An example in Figure 7.5 illustrates these cases.

To determine the quantitative value $\tau(v_i)$ in voxel v_i not only for individual tracts, but for an entire fiber bundle, all N tracts of the bundle that traverse v_i are taken into account. Then, the quantitative value $\tau(v_i)$ for v_i is the average of all tract-specific quantitative values $\tau_j(v_i)$ in v_i with $j = 1, \dots, N$:

$$\tau(v_i) = \frac{1}{N} \sum_{j=1}^N \tau_j(v_i). \quad (7.6)$$

7.2.5 Voxel-wise statistical analysis of individual fiber bundles

After spatial normalization, extraction of fiber bundles and gridding of quantitative values, statistical analysis is performed in *bundles of interest* (BOIs) to detect significant differences between groups of subjects. It is assumed that the extracted bundles of the datasets have been matched inter-individually to ensure that only anatomically corresponding bundles are used for the statistical analysis (see section 6.3). For each BOI, this results in a set of matched bundles, while each bundle belongs to a different dataset.

During the statistical analysis of a BOI, all voxels that belong to the BOI are analyzed. However, the matched bundles of the different datasets are not identical but differ in shape and extent. Voxels that are occupied in bundles of several datasets, may not be occupied in bundles of other datasets. As a result, the number of datasets that contribute to a voxel can vary. As size and extent of BOIs are known from the earlier bundling of fiber tracts (see section 7.2.3), contributions from other anatomical structures that do not belong to the BOI

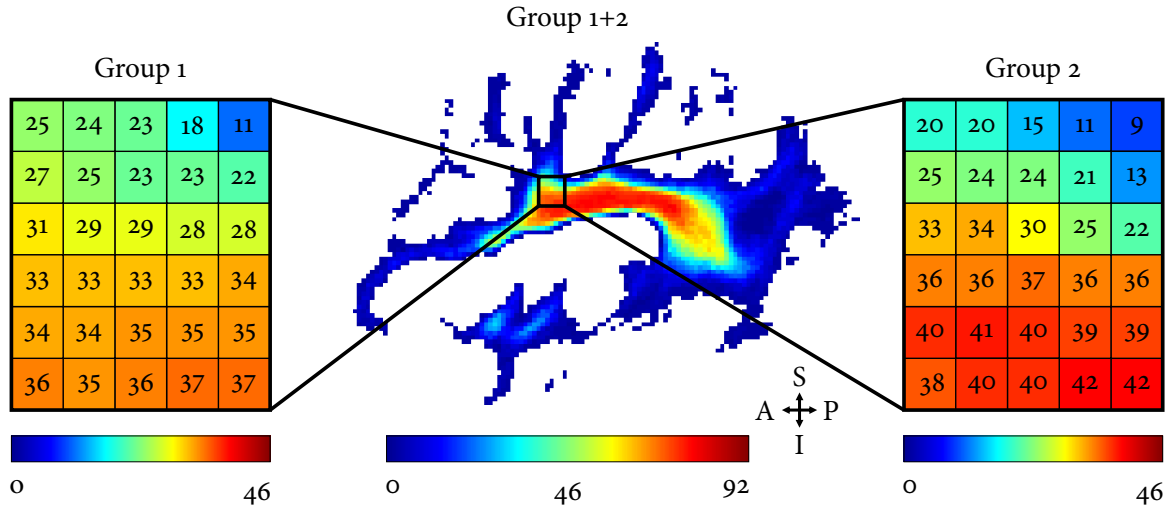


Figure 7.6: Number of datasets that contribute to the statistical analysis. For a central slice of the left superior longitudinal fasciculus (SLFt), the number of datasets that contribute to the group-based analysis is shown and encoded with pseudo colors. In the middle, the entire slice is displayed for both groups, each consisting of overall 46 datasets. On the left and the right, a smaller fragment provides a detailed view about the number of datasets per group that contribute to the independent statistical analysis in each voxel.

are suppressed. This benefits the statistical voxel-wise analysis of BOIs, as only those regions of a bundle from a dataset are considered that actually belong to the bundle. Figure 7.6 illustrates an example with different numbers of datasets that contribute to the statistical analysis of a selected bundle.

The partitioning of tracts into bundles and the subsequent matching enable the computation of probability maps that provide information about the number of datasets that contribute to each voxel of the BOIs (see section 6.2). With these reliability maps, regions of the BOI can be excluded from the analysis if too few datasets contribute to the region.

The statistical analysis of a BOI is finally performed for each voxel independently. On the basis of the quantitative values in each group, appropriate statistical tests (e.g., *Student's t-test*, *permutation test* (Nicholson, 2001)) are performed to determine statistical significance between groups. During the analysis of each BOI one test per voxel is performed, which requires adequate correction procedures to control statistical errors in such a multiple hypothesis testing scenario. As pointed out by Abdi (2007), the *Bonferroni* or *Šidák correction* are overly conservative for brain imaging data. Instead, the *false discovery rate* (FDR) as introduced by Benjamini and Hochberg (1995) represents a more appropriate approach to deal with multiple hypothesis testing.

7.3 Experiments

Various experiments were conducted to study the performance of the presented method and to investigate the influence of certain parameters on the analysis. For the experiments, the

Table 7.1: Parameters that were used to investigate the influence of the number of tracts, the distance between tract points and the gridding resolution on projecting and gridding of quantitative values.

Resolution in mm ³	Number of tracts	Distance between tract points in mm
0.3 × 0.3 × 0.3, 0.4 × 0.4 × 0.4,	300 000, 280 000,	0.3, 0.4, 0.6, 0.8, 1.0,
0.5 × 0.5 × 0.5, 0.6 × 0.6 × 0.6,	260 000, 240 000,	1.2, 1.4, 1.6, 1.8, 2.0,
0.7 × 0.7 × 0.7, 0.8 × 0.8 × 0.8,	220 000, 200 000,	2.2, 2.4, 2.6, 2.8, 3.0,
0.9 × 0.9 × 0.9, 1.0 × 1.0 × 1.0,	180 000, 160 000,	3.2, 3.4, 3.6, 3.8, 4.0,
1.1 × 1.1 × 1.1, 1.25 × 1.25 × 1.25,	140 000, 120 000,	4.2, 4.4, 4.6, 4.8, 5.0,
1.0 × 1.0 × 2.0, 1.5 × 1.5 × 1.5,	100 000, 80 000,	5.2, 5.4, 5.6, 5.8, 6.0,
1.25 × 1.25 × 2.5, 1.6 × 1.6 × 1.6,	60 000, 40 000, 20 000,	6.2, 6.4, 6.6, 6.8, 7, 8,
1.75 × 1.75 × 1.75, 2.0 × 2.0 × 2.0,	10 000, 8000, 7000,	9, 10, 11, 12, 13, 14, 15,
2.25 × 2.25 × 2.25, 2.5 × 2.5 × 2.5,	6000, 5000, 4000,	16, 17, 18, 19, 20, 21,
3.0 × 3.0 × 3.0, 3.5 × 3.5 × 3.5,	3000, 2000, 1000, 900,	22, 23, 24, 25, 26, 27,
4.0 × 4.0 × 4.0, 4.5 × 4.5 × 4.5,	800, 700, 600, 500	28, 29, 30
5.0 × 5.0 × 5.0		

same data as in section 6.1 was used. As certain processing steps differ depending on the experiment, each section summarizes relevant processing steps.

7.3.1 Dependency of the resolution, number of tracts and distances between tract points

Projection and gridding of quantitative values result in errors that are affected by the number of tracts, the distance between tract points and the gridding resolution. To prevent misrepresentation of quantitative values due to inadequate parameters, the influence of these parameters was investigated by conducting various experiments and comparing the results to previously generated ground truth datasets.

The latter datasets were created from a tractography dataset with 300 000 tracts and an average tract point distance of approximately 1.05 ± 0.17 mm (see section 6.1). Tracts were resampled to consist of equally distributed tract points with an average tract point distance of 0.3 mm. The dataset was not spatially normalized and resided in its native space. The *fractional anisotropy* (FA) was projected onto the tracts and attached values were normalized to a maximum value of 1. To investigate the influence of the parameters at different resolutions, tracts were rasterized and the attached FA values were gridded to a regular grid. This resulted in a set of ground truth volumes with different resolutions ranging from $0.3 \times 0.3 \times 0.3$ mm³ to $5 \times 5 \times 5$ mm³ (see Table 7.1, first row).

The dependency between number of tracts and resolution was studied by continuously reducing the number of tracts from 300 000 down to 500 using random sampling (see Table 7.1, second row). To investigate the influence of the distance between tract points at various resolutions, the tract point distance was successively increased from 0.3 mm to 30 mm (see Ta-

ble 7.1, third row). As this resampling of tracts resulted in the computation of new tract points, the FA values were projected onto the tract points after resampling. Hence, normalization of FA values was also performed after resampling. For both experiments, i.e., tract number reduction and tract resampling, tracts were rasterized and FA values that were attached to the tract points were gridded with the same resolutions as the corresponding ground truth volumes (see Table 7.1, first row). If multiple tracts traversed identical voxels their FA values were averaged (see the gridding procedure in section 7.2.4).

To investigate the effects of different number of tracts and varying tract point distances at various levels of resolution, the obtained FA maps were compared to the ground truth volumes by computing the average difference of quantitative values per voxel. If \mathbb{G} is a ground truth dataset and \mathbb{D} one experimental dataset, both having the same resolution, the average voxel difference $\Delta(\mathbb{G}, \mathbb{D})$ is given by:

$$\Delta(\mathbb{G}, \mathbb{D}) = \frac{1}{N} \sum_{i=1}^N \Delta_{\mathbb{G}, \mathbb{D}}(\mathbf{v}_i), \quad (7.7)$$

where N denotes the number of occupied voxels and $\Delta_{\mathbb{G}, \mathbb{D}}(\mathbf{v}_i)$ the difference between \mathbb{G} and \mathbb{D} in voxel \mathbf{v}_i that is occupied by tracts of \mathbb{G} or \mathbb{D} . As the quantitative values have been normalized and are bounded by $[0, 1]$, the maximum possible difference $\Delta_{\mathbb{G}, \mathbb{D}}(\mathbf{v}_i)$ in a voxel is 1. $\Delta(\mathbb{G}, \mathbb{D})$ therefore also ranges from 0 to 1.

The difference $\Delta_{\mathbb{G}, \mathbb{D}}(\mathbf{v}_i)$ is the difference of quantitative values between \mathbb{G} and \mathbb{D} in voxel \mathbf{v}_i . Thus, one has to take into account whether voxel \mathbf{v}_i is occupied by tracts from both datasets \mathbb{G} and \mathbb{D} or only by tracts from the ground truth dataset \mathbb{G} . If tracts of both datasets traverse \mathbf{v}_i (the tract densities $\rho_{\mathbb{G}}(\mathbf{v}_i)$ and $\rho_{\mathbb{D}}(\mathbf{v}_i)$ are higher than 0), $\Delta_{\mathbb{G}, \mathbb{D}}(\mathbf{v}_i)$ corresponds to the arithmetic difference between the quantitative values $\tau_{\mathbb{G}}(\mathbf{v}_i)$, $\tau_{\mathbb{D}}(\mathbf{v}_i)$ of \mathbb{G} and \mathbb{D} in \mathbf{v}_i . However, if only tracts of \mathbb{G} traverse \mathbf{v}_i but no tracts of \mathbb{D} , the difference between their respective quantitative values cannot be assessed in \mathbf{v}_i . This case can be considered the worst case scenario, in which the error is maximal. Consequently, the difference $\Delta_{\mathbb{G}, \mathbb{D}}(\mathbf{v}_i)$ is set to the maximum value of 1. Mathematically, $\Delta_{\mathbb{G}, \mathbb{D}}(\mathbf{v}_i)$ in voxel \mathbf{v}_i is given by:

$$\Delta_{\mathbb{G}, \mathbb{D}}(\mathbf{v}_i) = \begin{cases} \text{abs}(\tau_{\mathbb{G}}(\mathbf{v}_i) - \tau_{\mathbb{D}}(\mathbf{v}_i)) & \text{if } \rho_{\mathbb{G}}(\mathbf{v}_i) > 0 \wedge \rho_{\mathbb{D}}(\mathbf{v}_i) > 0, \\ 1 & \text{if } \rho_{\mathbb{G}}(\mathbf{v}_i) = 0 \oplus \rho_{\mathbb{D}}(\mathbf{v}_i) = 0. \end{cases} \quad (7.8)$$

7.3.2 Investigation of normalization-induced interpolation

The spatial normalization results in interpolation of quantitative values at tissue boundaries, which can influence the analyses. To investigate the prospects of the proposed technique to reduce this normalization-induced interpolation, the following experiments were conducted.

A first experiment was performed to confirm that individual fiber bundle-based processing leads to correct separation of quantitative values from fiber bundles. For this purpose, a new dataset was generated that consisted only of two partially overlapping fiber bundles – the inferior fronto-occipital fasciculus (IFO) and the uncinate fasciculus (UNC) (see Figure 7.7). Both bundles were extracted from a previously acquired, spatially normalized dataset that was clustered with the atlas-guided CATSER clustering and the *Hausdorff distance* (HD).

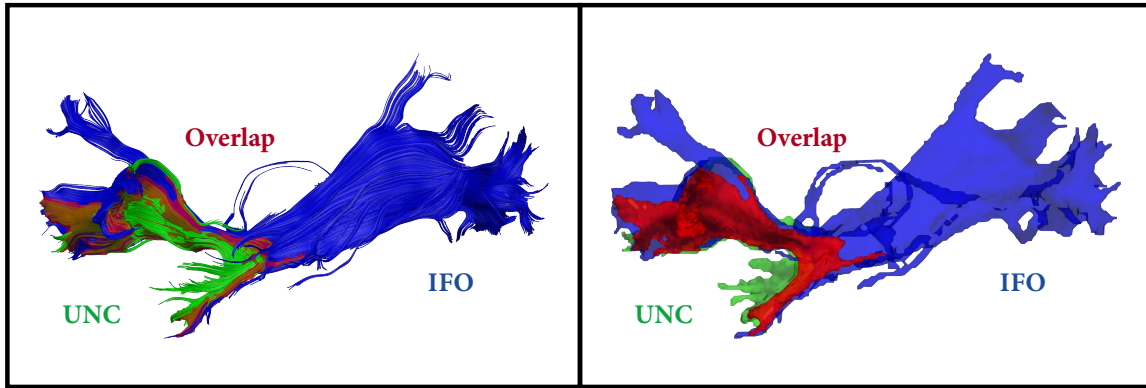


Figure 7.7: Tracts of the inferior fronto-occipital fasciculus – IFO and uncinate fasciculus – UNC (left image). The gridded voxels ($1 \times 1 \times 1 \text{ mm}^3$ resolution) that are occupied by the bundles are shown in the right image. In both images, the IFO is visualized in blue, the UNC in green and overlapping regions of both bundles are emphasized in red.

Acquisition, normalization and clustering parameters are specified in sections 6.1 and 6.4.1. Conventional gridding of the entire dataset and separate bundle-wise gridding of the individual bundles was performed to compare interpolation effects between both methods. With conventional gridding, the values that are attached to dataset are gridded for the entire dataset and contributions of different bundles are not handled individually. With bundle-wise gridding, values attached to the bundles of the dataset are gridded separately, respecting the previous division of tracts into different fiber bundles.

To investigate the interpolation in a controlled way, known values were attached to the tracts of both bundles. For the IFO a value of 1 and for the UNC a value of 2 was attached. Next, the newly generated dataset was rasterized and attached values were gridded to a 3D volume with conventional and bundle-based gridding. On the basis of the resulting 3D volumes, the distributions of gridded values were determined. As the gridding resolution influences the extent of the bundle overlap and thus influences the results, the experiment was performed for two different resolutions. While the first resolution corresponds to the resolution of the dataset in its native space ($1.25 \times 1.25 \times 2.5 \text{ mm}^3$), the second resolution was chosen to correspond to the resolution of the data in the MNI152 space of the atlas, in which the data resided after spatial normalization ($1 \times 1 \times 1 \text{ mm}^3$ isotropic resolution).

In a second experiment, normalization-induced differences between conventional and fiber bundle-based gridding were investigated for the previously acquired 46 real datasets (see section 6.1). FA values were projected onto the fiber tracts of the 46 datasets prior to spatial normalization (FA_{pre}). Then, fiber tracts and quantitative FA maps were transferred to the common template space and FA values were again projected onto the fiber tracts (FA_{post}). By performing atlas-guided clustering with CATSER and the CD similarity measure, the grouping of fiber tracts into bundles was achieved. Relevant parameters for fiber tracking, normalization and clustering are specified in sections 6.1 and 6.4.1. The fiber tracking datasets were rasterized to a grid with a resolution of $1 \times 1 \times 1 \text{ mm}^3$ and attached FA values (FA_{pre} and FA_{post}) were gridded with conventional and fiber bundle-based gridding. With

conventional gridding, one volume for FA_{pre} and one volume for FA_{post} was obtained per dataset. With bundle-based gridding, two volumes per fiber bundle were obtained for each dataset (again, one volume for FA_{pre} and one volume for FA_{post}). In overlapping regions of bundles, FA differences between conventional and bundle-based gridding were computed to investigate discrepancies between both gridding methods. Differences were thereby assessed independently for FA_{pre} and FA_{post} . The computed voxel value differences from all datasets were finally used to investigate interpolation-related differences between FA_{pre} and FA_{post} .

7.4 Results

7.4.1 Dependency of the resolution, number of tracts and distances between tract points

As described in section 7.3.1, the number of fiber tracts was gradually reduced from 300 000 to 500. Resulting tractography datasets were rasterized to various resolutions and average differences per voxel between gridded quantitative values of the experimental and the ground truth datasets were computed (see Eqs. 7.7 and 7.8). The results are presented in Figure 7.8. Variation of the number of tracts are shown on the y-axis and different resolutions on the x-axis. Resolutions are sorted according to their corresponding voxel volume from small to large voxels. The average differences per voxel are shown with pseudo colors. By inspecting the computed differences between experimental and ground truth datasets in Figure 7.8, it is evident that the average voxel differences depend on the number of fiber tracts and resolution. The observed differences between experimental and ground truth datasets ranged between 0 and 0.99. With decreasing number of tracts, fewer tracts were available in occupied voxels for the computation of quantitative values. When the number of tracts became too low, this resulted in voxels with incorrect values or voxels that were not occupied, which in return produced a higher average error per voxel. Especially in regions that contained few tracts, the average difference increased. To obtain correct quantitative values with the bundle-based projection-gridding approach, an adequate number of fiber tracts is necessary. Essentially, the higher the resolution the more tracts are necessary. At very high resolutions (e.g., $\leq 0.4 \text{ mm}^3$ isotropic resolution), a large number of tracts was mandatory while considerably fewer fiber tracts sufficed for high-moderate resolutions (e.g., $\geq 0.7 \text{ mm}^3$ isotropic resolution). In order to choose an adequate number of tracts for this projection-gridding approach one has also to consider that tracts are later grouped into fiber bundles and that outlier elimination will most certainly reduce the number of tracts in the final dataset. As both, the grouping and the outlier elimination, influence the gridded quantitative values, it is advised to use more tracts than necessary to minimize potential alteration of quantitative values during gridding (e.g., 10-20 % additional tracts).

For the second experiment, the distance between tract points was increased from 0.3 mm to 30 mm, which implicitly resulted in a reduction of tract points. Attached quantitative values were gridded to different resolutions and the average difference of values per voxel was computed between obtained experimental and ground truth datasets (see Eqs. 7.7 and 7.8). The results are depicted in Figure 7.9. The average tract point distances are shown on the

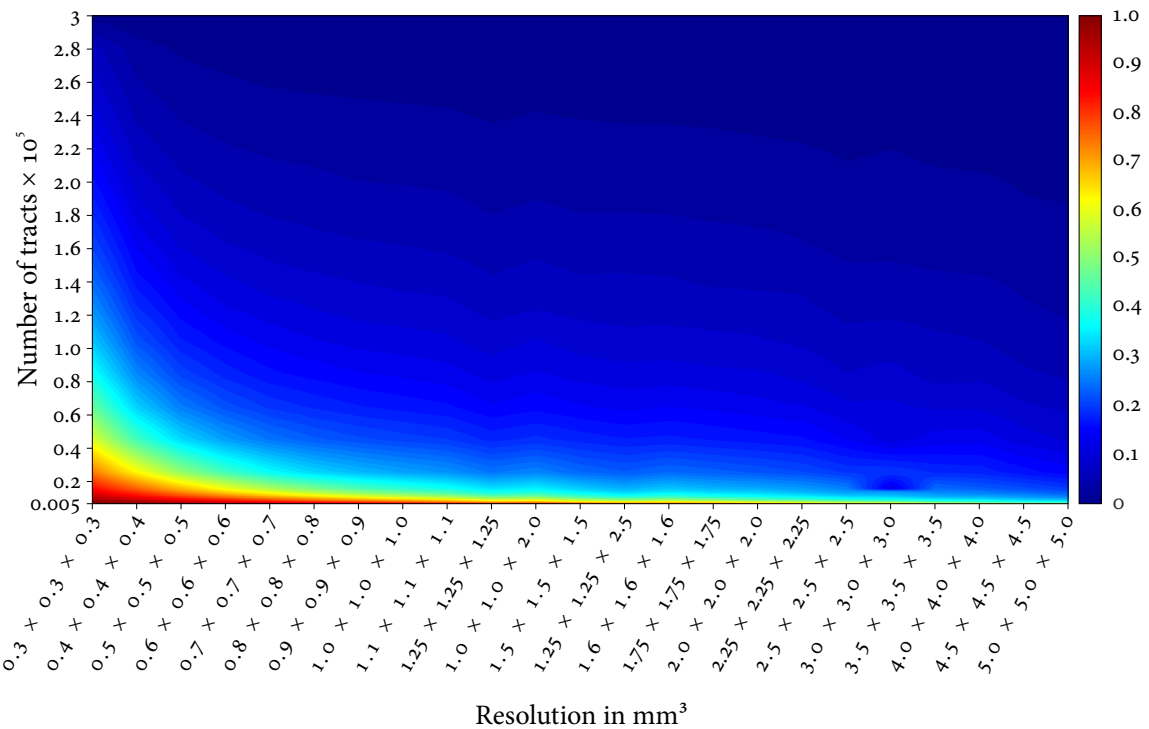


Figure 7.8: Average difference per voxel in dependence of resolution and number of tracts. Quantitative values have been projected onto a high resolution dataset. The number of tracts (y-axis) was gradually reduced from 300 000 to 500 and remaining tracts were gridded to different resolutions (x-axis). The average differences of quantitative values between experimental and ground truth datasets in occupied voxels are shown in pseudo colors.

y-axis and different resolutions on the x-axis. For all resolutions, increased distances between tract points resulted in elevated differences of quantitative values between experimental and ground truth datasets. Compared to the previous experiment, the maximum observed differences were lower and ranged only between 0 and 0.55. While increased tract point distances altered the course of the fiber tracts, the projection of quantitative values became less reliable as values were projected onto fewer tract points. If the native resolution was too high and the tract point distances were too large (e.g., 0.4 mm³ isotropic resolution and a tract point distance of 10 mm), various voxels were insufficiently taken into account during the projection of quantitative values onto the tracts. Due to this inaccurate projection, incorrect values were gridded. For larger tract point distances (5 mm and larger), the average differences were more susceptible to variations in resolution. Basically, the higher the resolution, the shorter distances between tract points have to be, to preserve quantitative values correctly. It is therefore necessary to use sufficiently small tract point distances to preserve quantitative values for the analysis. In general, it is advised to resample tracts to an average tract point distance of at least 1 mm, except for very high resolutions (e.g., ≤ 0.4 mm³ isotropic resolution) where the tract point distance should be adequately adjusted to minimize alterations of quantitative values (see Figure 7.9).

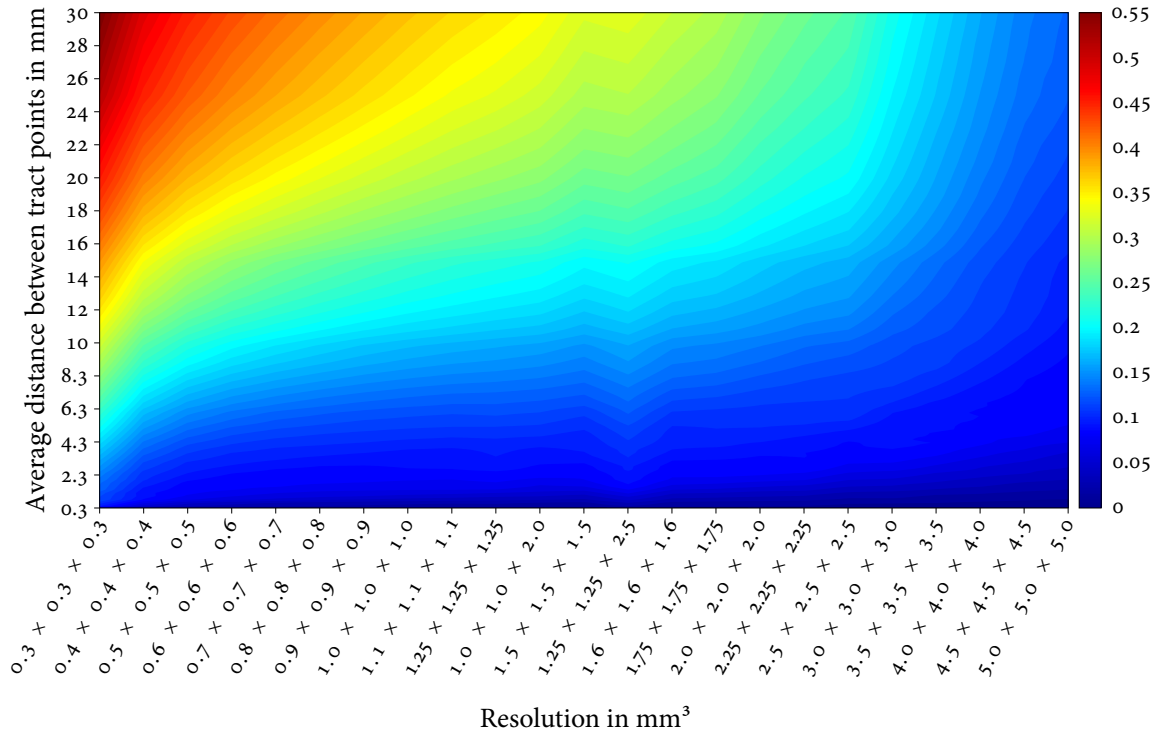
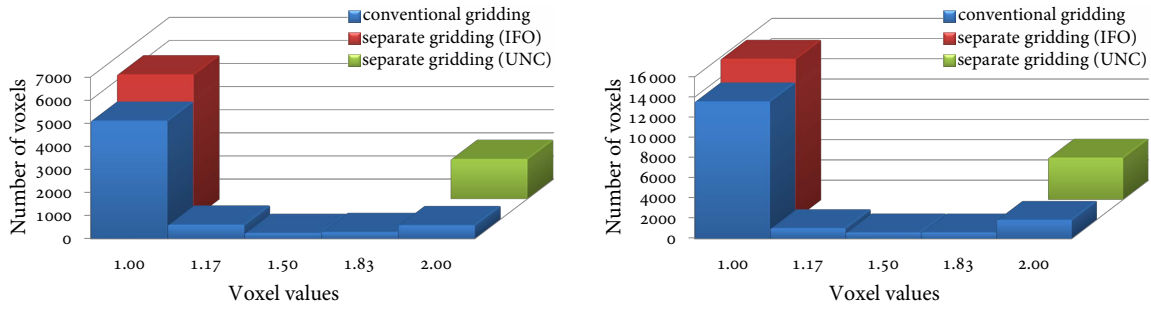


Figure 7.9: Average difference per voxel in dependence of resolution and tract point distance. The average distance between tract points (y-axis) was gradually increased from 0.3 mm to 30 mm. Quantitative values were projected onto obtained datasets and tracts were gridded to different resolutions (x-axis). The average differences of quantitative values between experimental and ground truth datasets in occupied voxels are shown in pseudo colors.

7.4.2 Investigation of normalization-induced interpolation

To confirm that the bundle-based approach facilitates the separation of contributions from different fiber bundles, a first experiment was conducted to compare conventional and bundle-based gridding of values (see section 7.3.2). For this purpose a reduced dataset was generated that consisted only of two overlapping fiber bundles (IFO and UNC). Different quantitative values were attached to the bundles (IFO = 1, UNC = 2), and the resulting dataset was gridded to a regular grid. First, conventional gridding was used to grid values of all tracts irrespective of the bundle-membership of the tracts. Second, the gridding was performed separately, by gridding the tracts of each bundle independently. The gridding was performed for two resolutions ($1.25 \times 1.25 \times 2.5 \text{ mm}^3$ and $1 \times 1 \times 1 \text{ mm}^3$) and the resulting distributions of values in the generated 3D volumes were investigated.

In Figures 7.10a and 7.10b the distribution of gridded values is visualized for both resolutions. The distributions are shown with histograms for conventional gridding (first row in blue), for separate gridding of the IFO (second row in red) and for separate gridding of the UNC (third row in green). With conventional gridding (first row), all tracts were gridded irrespective of the fiber bundle they belong to. In overlapping areas of the bundles, this resulted in interpolation of bundle-specific values (cmp. Figure 7.7). As the gridding takes into account the



(a) Gridding at $1.25 \times 1.25 \times 2.5 \text{ mm}^3$ resolution.

(b) Gridding at $1 \times 1 \times 1 \text{ mm}^3$ resolution.

Figure 7.10: Distribution of values after gridding of fiber bundles with different methods. A single dataset that consisted of two overlapping fiber bundles – the inferior fronto-occipital fasciculus (IFO) and the uncinate fasciculus (UNC) – was gridded with conventional and individual bundle-based gridding at two different resolutions (a), (b). By performing separate, bundle-based gridding (second and third rows in red and green), values that were attached to the bundles (IFO = 1, UNC = 2) are preserved. With conventional gridding, interpolation between values occurs in overlapping regions (first row in blue).

number of tracts that traverse through each voxel, a continuous interpolation of quantitative values occurred in overlapping regions. If bundles are gridded individually (second and third row in Figures 7.10a, 7.10b), two separate 3D volumes are obtained, each consisting of one compartment that contains the preserved values of the respective bundles (IFO = 1, UNC = 2). No interpolation occurred and overlapping regions in the resulting volumes were not corrupted by values of the other bundle.

While these observations are consistent for both resolutions, differences were observed in the proportion of interpolated voxels with conventional gridding. With decreasing resolution, the voxel size as well as the proportion of interpolated voxels increases. With a resolution of $1.25 \times 1.25 \times 2.5 \text{ mm}^3$ (Figure 7.10a), the UNC consisted of 1728 voxels whereas the IFO consisted of 6251 voxels. With a total of 1143 overlapping voxels, approximately 66 % of the UNC voxels and 18 % of the IFO voxels were subject to interpolation. Hence, attached values were preserved for only 34 % of the UNC and for 82 % of the IFO voxels. With a gridding resolution of $1 \times 1 \times 1 \text{ mm}^3$ (Figure 7.10b), the separation of values in the overlapping regions of the bundles improved. For the increased resolution, the UNC consisted of 4149 and the IFO of 15 863 voxels. A total of 2283 voxels were overlapping, which corresponded to 55 % voxels of the UNC and 14 % for the IFO. Compared to the resolution of $1.25 \times 1.25 \times 2.5 \text{ mm}^3$ (see above), more voxel values were preserved by percentage (45 % for the UNC and 86 % for the IFO). In principle, this implies that the separate gridding especially aids fiber bundles with a high proportion of overlapping voxels.

In the second experiment, differences between conventional and fiber bundle-based gridding were investigated for quantitative FA values that were attached to previously acquired datasets before (FA_{pre}) and after (FA_{post}) spatial normalization. After normalization, the tracts of the individual datasets were grouped into fiber bundles with atlas-guided clustering. For FA_{pre} and FA_{post} , conventional and separate gridding of fiber bundles was performed and dif-

ferences of quantitative values between both gridding methods were computed in overlapping voxels of fiber bundles (separately for FA_{pre} and FA_{post}).

Resulting distributions of differences between FA_{pre} and FA_{post} are presented in Figure 7.11 for all fiber bundles. Histograms are shown for 100 discrete intervals. Identical tracts, bundles and regions were analyzed for both FA_{pre} and FA_{post} . Distributions for FA_{pre} and FA_{post} in the histograms are therefore based on the same number of voxels.

Closer inspection of the distribution of the obtained voxel value differences revealed diverse effects. Investigating the distribution differences for FA_{post} – the FA values that were attached after spatial normalization (depicted with red bars) – revealed that conventional gridding and independent gridding of individual fiber bundles led to differences in the FAs of the bundles. This can be seen in Figure 7.11 where the height of the red bars denotes the number of voxels with FA_{post} differences between conventional gridding and independent gridding. If conventional gridding and independent gridding would have been identical, the voxel value differences would have been zero for all analyzed voxels and only one single red bar with a voxel value difference of 0 would have existed. Normalization-induced interpolation can be ruled out as a reason for these differences, as FA values were attached to the tracts after spatial normalization. Since identical datasets were used, the observed discrepancies are thus directly related to differences in the gridding methods. With conventional gridding all tracts of a dataset were gridded, irrespective of the bundle the tracts belong to. In contrast, for bundle-based gridding, each bundle was gridded independently and only those tracts were gridded that belonged to the same bundle. These different processing strategies influenced the final voxel value. If only one fiber bundle contributed to a voxel, conventional and bundle-based gridding led to identical results and the same voxel value. However, if N fiber bundles overlapped and occupied the same voxel (with $N \geq 2$), differences between the gridding methods arose for this voxel. By applying conventional gridding in a case where $N \geq 2$, the tracts of all N bundles were used to determine the final value in the voxel. For bundle-based gridding, however, each bundle was gridded separately. Hence, the gridding was performed N times. Each time a single bundle was gridded, only those tracts were used for the gridding that belonged to the actual bundle. As the tracts of the other $N - 1$ bundles were neglected, they contributed nothing to the final voxel value. The slight differences between conventional and bundle-based gridding, in voxels that were occupied by multiple bundles, were ultimately due to variations in the tracts that were gridded with conventional and bundle-based gridding.

Besides these gridding-related alterations, discrepancies were also observed for quantitative FA values that were attached before (FA_{pre}) and after (FA_{post}) spatial normalization (see the differences between the blue (FA_{pre}) and the red (FA_{post}) bars in Figure 7.11). For the gridding of the attached quantitative values for FA_{pre} and FA_{post} , identical tracts and fiber bundles were used. Hence, the contributions of fiber bundles to the voxels during the gridding were identical for FA_{pre} and FA_{post} . This implies that the observed differences in the distributions between FA_{pre} and FA_{post} (blue and the red bars in Figure 7.11) can not be explained with the previously described gridding-related differences (see previous paragraph). As the additional processing steps for FA_{pre} and FA_{post} were also identical, the only plausible explanation for the discrepancies in the distributions between FA_{pre} and FA_{post} are normalization-induced interpolation differences that occurred during the gridding of fiber tracts with conventional gridding.

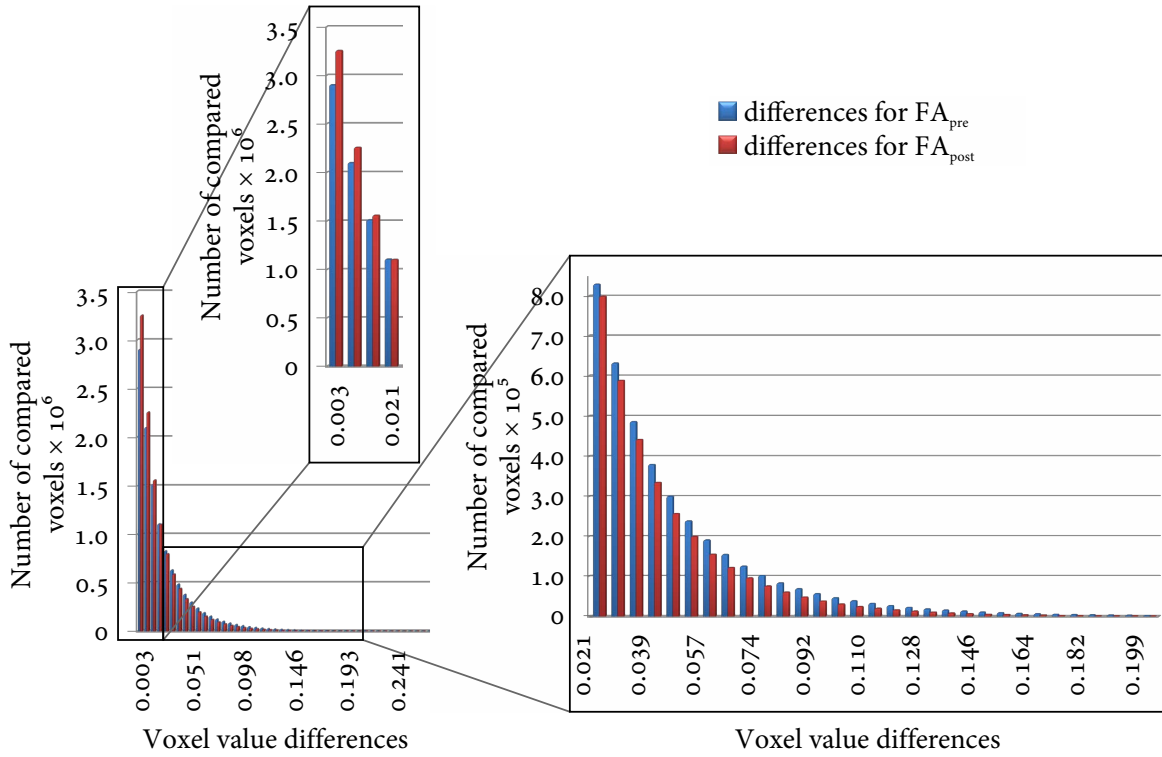


Figure 7.11: Distributions of voxel value differences between conventional and fiber bundle-based gridding for quantitative FA values attached before (FA_{pre} in blue) and after (FA_{post} in red) spatial normalization.

A more detailed investigation of the distribution differences between FA_{pre} and FA_{post} supports this hypothesis. The histogram in Figure 7.11 can be divided into two parts. In the left part of the histogram, the number of voxels for FA_{post} exceed the number of voxels for FA_{pre} (the red bars are higher than the blue bars in Figure 7.11). Strictly spoken, compared to FA_{pre} more voxels of FA_{post} have lower voxel value differences. In the right part of the histogram this effect is inverted. Here, more voxels of FA_{pre} have higher voxel value differences in comparison to FA_{post} (the blue bars are higher than the red bars in Figure 7.11). Overall, the observed voxel value differences were higher for FA_{pre} than for FA_{post} . Practically, this means that the gridding differences between conventional and bundle-based gridding were higher for FA_{pre} than for FA_{post} , which is a strong indicator that the bundle-based gridding preserved the quantitative values. For FA_{post} , quantitative values were attached after spatial normalization. The interpolation of quantitative values had thus already occurred and the observed voxel value differences can be attributed to the gridding methods only (see the earlier paragraph on gridding-related alterations). In contrast, for FA_{pre} quantitative values were attached to the tracts before spatial normalization. Due to the utilization of fiber tracts, no interpolation of quantitative values took place during this normalization step (see section 7.2.3). As demonstrated with the first experiment in this section, bundle-based gridding preserves the attached quantitative values and avoids corruption of the values from other fiber bundles. For conventional gridding, however, attached values are not preserved as all tracts are gridded

irrespective of the bundle they belong to. Hence, for FA_{pre} and conventional gridding, the interpolation of the values was only postponed and occurred later during the conventional gridding. As the attached values are preserved with bundle-based gridding but are interpolated with conventional gridding, observed differences between conventional and bundle-based gridding were the result of the normalization-induced interpolation and gridding-related alterations. However, due to the identical processing of FA_{pre} and FA_{post} , gridding-related alterations were identical in both cases. As they can thus be ruled out, spatial normalization remains as the source for the observed differences between FA_{pre} and FA_{post} .

8

Chapter 8

Applications

The beginning of knowledge is the discovery of something we do not understand.

— Frank Herbert (1920–1986)

In order to investigate the applicability of the presented quantitative analysis method and to determine whether it is capable to detect statistically significant alterations in white matter diffusion, two studies were conducted. Initially, diffusion-associated hemispheric differences in white matter bundles were studied for a group of healthy volunteers. Subsequently, hemispheric differences of diffusion were also investigated for selected fiber bundles of schizophrenic patients. Findings were finally compared to the previously obtained results for healthy volunteers.

8.1 Assessing hemispheric white matter differences in healthy volunteers

8.1.1 Introduction

The brain as a whole consists of numerous structures and is divided into two hemispheres that are partly characterized by a profound hemispheric asymmetry and bilateralization in both structure and function (Hellige, 1993). The earliest observations of brain asymmetry were described by Broca (1865) and Wernicke (1874), who discovered the specialization of the left hemisphere for speech and language processing and the impairment in speech synthesis, language comprehension and syntactic processing that resulted from localized injuries of the left hemisphere (Toga and Thompson, 2003; Hugdahl, 2005).

These initial observations marked the beginning of a new research area that aims at unraveling differences in the structural and functional organization of the brain hemispheres. Sperry (1961), for example, performed split-brain experiments and investigated epileptic patients who had undergone a *corpus callosotomy* in which the inter-hemispheric connections

of the corpus callosum were surgically severed¹. His split-brain experiments opened a new window to study brain function and facilitated the investigation of functional differences between the hemispheres. The independent functional analysis of the cerebral hemispheres provided the unique opportunity to study inter-hemispheric communication and interaction of processing centers. While Sperry's experiments provided insight into a variety of neural mechanisms (Gazzaniga, 2005), investigations of other scientists, such as Galaburda et al. (1978) and Geschwind (1979), shed light on inter-hemispheric differences at a structural and cytoarchitectural level, which later culminated in their theory of cerebral lateralization (Geschwind and Galaburda, 1987; Toga and Thompson, 2003).

Hemispheric asymmetry and bilateralization of the brain are characteristics that are not unique to humans, but are in fact a ubiquitous phenomenon that exists in other species (Güntürkün, 2005) and can be observed even in simple multicellular organisms such as *Caenorhabditis elegans* (see section 1.1 Origin of the brain; Corballis (2009); Goldsmith (2011)). Because of the widespread occurrence of cerebral asymmetry among species, it is believed that genetic factors influence the expression of asymmetries, though the underlying genetic mechanisms still remain elusive (Corballis, 2009). In addition to genetic factors, research indicates that cerebral asymmetries are at least modulated by environmental and biological factors (Hellige, 1993). Due to complex interactions with the environment, asymmetries are shaped across the life span (Hellige, 1993), but are also influenced as the brain matures (e.g., during adolescence, Geschwind and Galaburda (1987)).

Besides traditional experiments (e.g., split-brain, see above), MRI has been used as a complementary method to study cerebral lateralization. Conventional structural MRI techniques were primarily, but not exclusively, used to assess hemispheric differences in the neuronal processing centers of the grey matter (Good et al., 2001; Kovalev et al., 2003), while altered diffusion in white matter pathways was investigated with DWI (Fillard et al., 2003; Park et al., 2004; O'Donnell et al., 2009; Takao et al., 2011). In order to determine whether the presented method for quantitative analysis (see chapter 7) is capable to detect statistically significant alterations in white matter diffusion, hemispheric differences of selected fiber bundles were investigated for a group of healthy volunteers.

8.1.2 Data acquisition, data processing and statistical analysis

For the investigation of hemispheric differences in selected white matter tracts, the previously acquired 46 datasets of healthy volunteers were employed. Subsequently to the data acquisition, fiber tracking was performed and tracts were resampled to an average tract point distance of 1 mm. Acquisition and tractography parameters are described in section 6.1. For the quantitative analysis as described in chapter 7, quantitative values (FA, ADC, AD and RD; see section 2.3.3) were projected onto the fiber tracts (see section 7.2.2). In order to achieve the superposition between the bundles of opposite hemispheres, all 46 datasets were flipped from left to right. This resulted in a total of 92 datasets for both two groups – 46 original datasets and 46 flipped datasets.

¹A *corpus callosotomy* is a treatment for epilepsy introduced by Van Wagenen and Herren (1940) to prevent epileptic seizures from spreading across the cerebral hemispheres.

8.1 Assessing hemispheric white matter differences in healthy volunteers

Spatial normalization of all 92 datasets was performed with the ANTs framework using the FA maps (Klein et al., 2009). During this process, an FA template, displacement fields and affine transformation matrices were obtained for all datasets. Using this information, the fiber tracts of the individual datasets were transferred to the common template space. Normalization steps and parameters were identical to the steps and parameters in section 6.1.

In order to extract the fiber bundles, fully automated atlas-guided cluster analysis was performed for all 92 datasets using the *matched point distance* (MPD) and the previously generated white matter atlas (see Part II). Other clustering parameters corresponded to the parameters that were used previously (see section 6.1). To relate the clusters to the corresponding atlas classes, extracted fiber bundles were spatially matched to the atlas bundles (see section 6.3). As hemispheres are symmetrical, only nine *bundles of interest* (BOIs) were obtained for each dataset: forceps major (Fmaj), the frontal projection of the corpus callosum (forceps minor – Fmin), anterior thalamic radiation (ATR), gyrus part of the cingulum cingulate (CGC), hippocampal part of the cingulum (CGH), corticospinal tract (CST), inferior fronto-occipital fasciculus (IFO), temporal part of the superior longitudinal fasciculus (SLFt) and the uncinate fasciculus (UNC).

To conduct the voxel-wise quantitative analysis for each BOI, tracts were rasterized to a regularly spaced grid with $1 \times 1 \times 1 \text{ mm}^3$ resolution and attached quantitative values were investigated separately for each BOI. For each dataset and each diffusion property this resulted in nine volumes (one for each BOI) that were used for the quantitative analysis. For the statistical analysis, *permutation tests* with 1000 permutations (Nicholson, 2001) were performed in the voxels of each BOI, as long as at least 10 datasets (out of 46) per group contributed data to the voxels. To correct for multiple comparisons the *false discovery rate* – FDR (Benjamini et al., 2006) was used with a significance level $p < 0.01$. Only regions with at least 50 connected voxels were deemed valid.

8.1.3 Results

The quantitative analysis of hemispheric differences of diffusion in selected white matter fiber bundles was successfully performed as described in the previous section. Statistically significant FA differences in corresponding bundles of the left and right hemisphere were observed. After applying the FDR correction, results were still statistically significant in certain regions of the CGC, UNC, CST, IFO and SLFt. In Figures 8.1–8.5, the fiber bundle regions that were statistically significantly different between both hemispheres as well as quantitative values in these regions are shown. In each figure, a volume rendering (on the left) displays the analyzed bundle and the spatial location of voxels that had statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences before the correction for multiple hypothesis testing are shown in green, voxels that were still significant after the FDR correction are colored in brown. To summarize the findings, quantitative values (FA, ADC, AD, RD) of the left and right hemisphere of all datasets in statistically significant voxels after FDR correction are shown with box plots. The centerline in the boxes denotes the median (second quartile). The bottom and the top of a box correspond to the first and the third quartile and define the *inter-quartile range* (IRQ). The lines that emerge from the top and bottom of a box contain all remaining data points that are in the

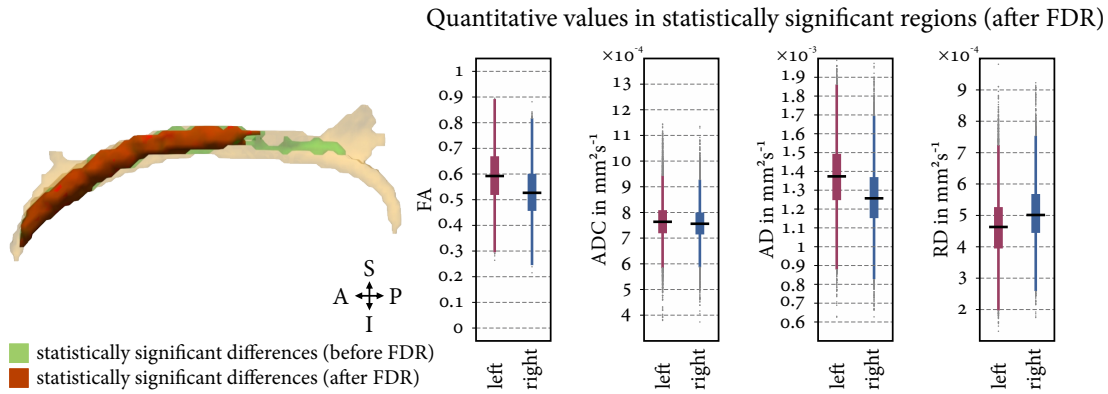


Figure 8.1: Hemispheric differences in the CGC of healthy volunteers. The volume rendering on the left shows the analyzed CGC and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant voxels (after FDR, $p < 0.01$).

range of $\pm 1.5 \times \text{IRQ}$. For the statistically significantly different regions in the bundles, Table 8.1 provides a qualitative overview of FA, ADC, AD and RD differences.

For the CGC, UNC, CST, IFO and SLFt prominent differences in FA, AD and RD values between the left and right hemisphere were found. Differences in ADC, which represents the average diffusion along all three orthogonal axes of the diffusion tensor (see sections 2.1 and 2.3.2), were often barely noticeable and may thus remained undetected. Observed differences in ADC and FA were the result of differences between eigenvalues, and thus differences in the *axial diffusivity* – AD (the first eigenvalue λ_1) and *radial diffusivity* – RD (the average of the second and third eigenvalue $(\lambda_2 + \lambda_3)/2$).

By investigating the quantitative values for each bundle in detail, it becomes evident that AD and RD influence FA and ADC in different ways. In the anterior part of the CGC for example (Figure 8.1), the FA of the left hemisphere is increased compared to the right hemisphere. Observed differences in the voxels are thereby the result of an AD increase and an RD decrease (left vs. right). Differences in the ADC, however, were minimal. For the UNC (Figure 8.2) differences in diffusion between the left and right UNC are reversed compared to the CGC. While AD in the left hemisphere was slightly decreased compared to the right hemisphere, RD was increased which resulted in a decreased FA and a slightly higher ADC (left vs. right). For the CST (Figure 8.3), FA was decreased as in the UNC (left vs. right).

Table 8.1: Summary of differences in diffusion parameters for statistically significant regions of bundles from the left and right hemisphere (healthy volunteers).

Bundle	Average quantitative values			
	FA	ADC	AD	RD
CGC	$l \gg r$	$l \approx r$	$l \gg r$	$l < r$
UNC	$l \ll r$	$l > r$	$l < r$	$l \gg r$
CST	$l \ll r$	$l > r$	$l \approx r$	$l \gg r$
IFO – R1	$l \ll r$	$l > r$	$l < r$	$l \gg r$
IFO – R2	$l \ll r$	$l > r$	$l < r$	$l \gg r$
SLFt – R1	$l < r$	$l > r$	$l < r$	$l \gg r$
SLFt – R2	$l \gg r$	$l \approx r$	$l \gg r$	$l \ll r$

8.1 Assessing hemispheric white matter differences in healthy volunteers

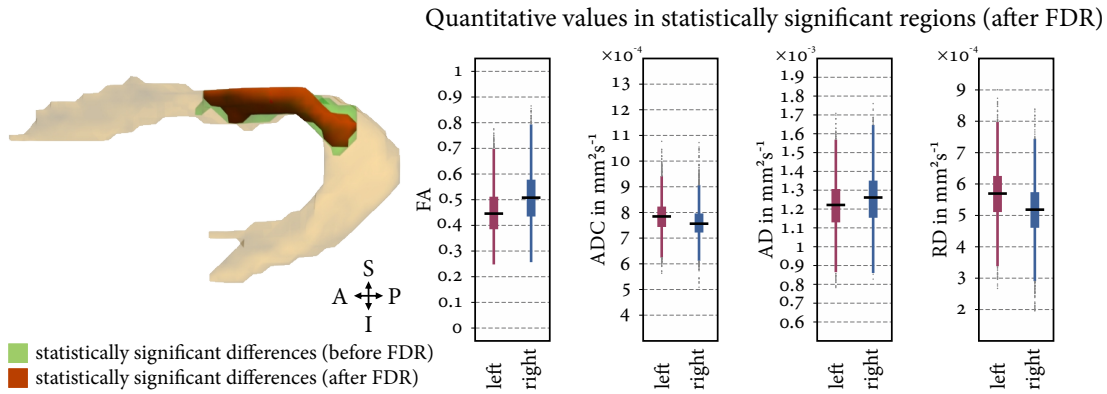


Figure 8.2: Hemispheric differences in the UNC of healthy volunteers. The volume rendering on the left shows the analyzed UNC and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant voxels (after FDR, $p < 0.01$).

However, in contrast to the UNC, these differences were primarily governed by an increased RD of the left CST. The AD of the left CST is nearly identical to the AD of the right hemisphere, which led to increased ADC values in the left CST region. For the IFO bundle (Figure 8.4) two regions show statistically significant differences between the left and right hemisphere. These differences are in line with the pattern of differences observed in the UNC (see above).

In both left IFO regions FA was decreased and ADC slightly increased compared to the right hemisphere. The differences in both IFO regions were due to decreased AD and increased RD of the left IFO (see UNC, above). Comparing quantitative values between both significantly different regions of the IFO, only minor, negligible variations were found.

The SLFt bundle also consists of two regions that are significantly different between the left and the right hemisphere (Figure 8.5). However, in contrast to the IFO, the differences between the regions were opposed, though the regions were spatially close. While FA in region R1 was decreased it was increased for region R2 (left vs. right). For region R1 the increased FA was the result of a slightly reduced AD and an increased RD in the left hemisphere compared to the right hemisphere. The ADC for the left hemispheric region was slightly increased. In region R2 the situation is different. In R2, hemispheric differences of the FA were the result of a major AD increase and a more pronounced RD decrease (left vs. right). The ADC is almost identical and differences were barely noticeable. Considering the opposed differences between regions R1 and R2 of the SLFt, the question arises where these differences originate. At a microstructural level, the SLFt (often termed arcuate fasciculus) is composed of three fiber segments (anterior, posterior and long) that interconnect Broca's, Geschwind's and Wernicke's territory (see Figure 8.6). By comparing the spatial location of R1 and R2 with the anatomically labeled image in Figure 8.6, it appears that region R1 is primarily located in the anterior segment connecting Broca's and Geschwind's territory, while region R2 is located in the long segment that ties Broca's to Wernicke's territory. Observed differences between R1 and R2 were most likely associated with differences in the distinct segments of the SLFt.

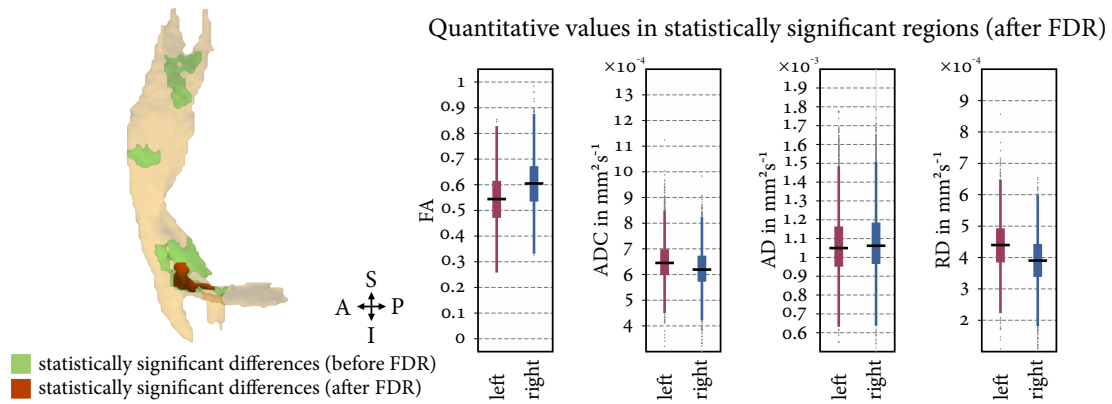


Figure 8.3: Hemispheric differences in the CST of healthy volunteers. The volume rendering on the left shows the analyzed CST and voxels that have statistically significant differently FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant voxels (after FDR, $p < 0.01$).

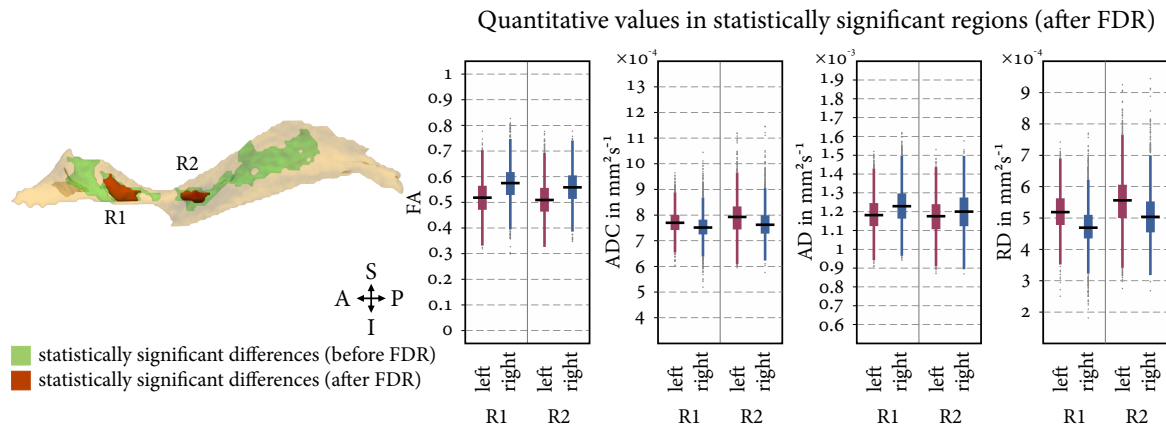


Figure 8.4: Hemispheric differences in the IFO of healthy volunteers. The volume rendering on the left shows the analyzed IFO and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant regions R1 and R2 (after FDR, $p < 0.01$).

8.1 Assessing hemispheric white matter differences in healthy volunteers

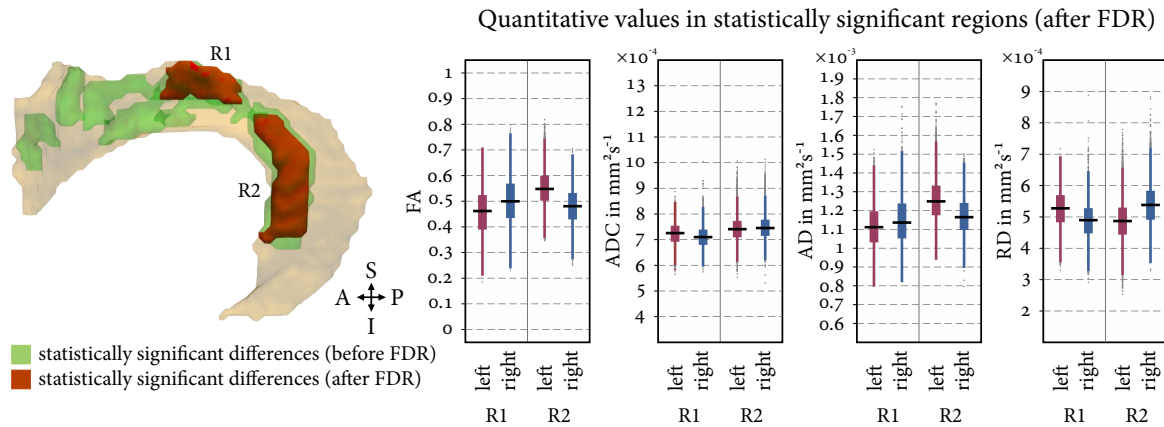


Figure 8.5: Hemispheric differences in the SLFt of healthy volunteers. The volume rendering on the left shows the analyzed SLFt and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant regions R1 and R2 (after FDR, $p < 0.01$).

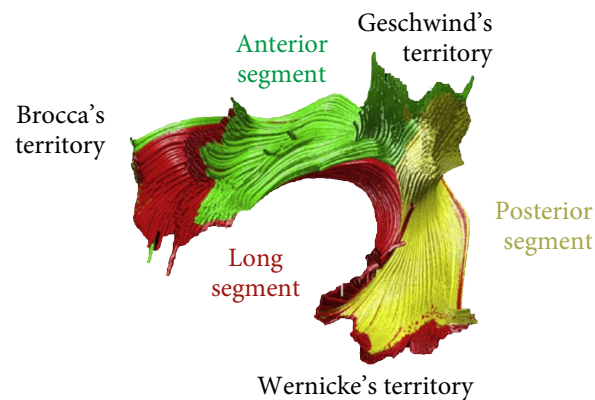


Figure 8.6: The segments of the arcuate fasciculus and the regions they connect. Taken and modified from Catani and Mesulam (2008).

8.2 Assessing hemispheric white matter differences in schizophrenic patients

8.2.1 Introduction

Schizophrenia is a complex mental disorder that is characterized by an impairment of thinking, cognition and behavior (Picchioni and Murray, 2007). While it is a multifactorial illness caused by genetic and environmental factors, the underlying cerebral mechanisms that finally lead to the typical psychopathological symptoms are not well understood (Picchioni and Murray, 2007; MacDonald and Schulz, 2009; Insel, 2010). With the advent of MRI, hundreds of studies have been conducted to explore the profound neuropathological changes that are associated with schizophrenia (van Os and Kapur, 2009). The comprehensive reviews by Shenton et al. (2001) and Kubicki et al. (2007) provide a broad overview of various MRI studies. Reported findings are diverse and suggest an involvement of a large number of cerebral structures that are scattered throughout the brain (Shenton et al., 2001). Due to the heterogeneity of findings and symptomatology, researchers still struggle in developing a comprehensive, generally accepted theory that explains all observations and neuropathological symptoms. Consequently, multiple complementary theories have evolved, trying to relate findings to, for example, neurodevelopmental issues (Weinberger, 1986; Owen et al., 2011), abnormal functions of either glutamate (Javitt, 2010) or dopamine receptors (Howes and Kapur, 2009) or dysfunctional connectivity between brain regions (Friston, 1998, 2002).

The latter theory, also termed dysconnection hypothesis, postulates that schizophrenia is the result of abnormal synaptic plasticity, impaired structural connectivity of white matter tracts or a combination of both (Stephan et al., 2009; Bullmore et al., 1997). These alterations might culminate into abnormal communication between functionally related, but spatially disparate brain regions (Stephan et al., 2009; Whitford et al., 2011). Experiments already provided evidence for disturbed synaptic function and plasticity (Stephan et al., 2009; Medkour et al., 2010; Guo et al., 2013). In theory, such abnormal structural connectivity might result in subtle changes of the underlying white matter microstructure, thereby influencing diffusion of molecules. However, studies that investigated altered diffusion with DWI remain inconclusive (Kubicki et al., 2007). While various studies observed altered diffusion (e.g., Ardekani et al. (2011)), other studies showed no abnormal diffusion in schizophrenic patients (for an overview of findings, see the studies by Kanaan et al. (2005), Kubicki et al. (2007) and Whitford et al. (2011)). These heterogeneous results only reinforce the fact that schizophrenia is a multifaceted disease with distinct manifestations.

To investigate the presence of potential hemispheric diffusion differences in selected white matter bundles of schizophrenic patients, a small study was conducted. Differences in corresponding bundles of the left and right hemisphere were investigated and findings were compared to the previously obtained results in healthy volunteers.

8.2.2 Data acquisition, data processing and statistical analysis

15 schizophrenic patients (4 female, 41 ± 11.1 years old; 11 male, 35 ± 9.0 years old) were included and measured on a clinical 3 T whole-body MR-Scanner (Tim Trio, Siemens Healthcare, Erlangen, Germany). All patients were diagnosed with schizophrenia according to DSM-IV (American Psychiatric Association, 2000). At the time point of the measurements, patients were medication-free or clinically stabilized by standard antipsychotic medication.

The study was approved by the Ethics Committee of the Jena University Hospital. Participants provided informed written consent to participate in this study in accordance to the statement of the Ethics Committee. To enable comparability between healthy volunteers and patients, the data acquisition and data processing was performed as previously described in section 8.1.2. All 15 datasets were flipped from left to right, which resulted in a total of 30 datasets for both two groups. The statistical analysis was performed for the same BOIs as in section 8.1.2. However, as only a small group of 15 patients was enrolled compared to the 46 healthy volunteers in section 8.1, parameters for the quantitative analysis were slightly adjusted. *Permutation tests* (1000 permutations) were performed in the voxels of each bundle, as long as at least five datasets per group contributed data to each voxel. To correct for multiple comparisons the FDR (Benjamini et al., 2006) was used with an adjusted significance level ($p < 0.05$). Only regions with at least 50 connected voxels were deemed valid.

8.2.3 Results

Using the datasets of schizophrenic patients, the quantitative analysis of hemispheric differences in diffusion was performed in the nine selected fiber bundles. As in section 8.1.3, statistically significant differences were observed between bundles of the left and right hemisphere. After adjusting the p-value using the FDR, differences in diffusion between regions of the left and right hemisphere were still statistically significant for the CGC, UNC, IFO and SLFt fiber bundle. In the CST, however, only few voxels showed statistically significant differences. For the sake of completeness, results for the CST are nevertheless presented (before FDR correction). In Figures 8.7–8.11, regions with statistically significant differences in diffusion between the left and right hemisphere are shown for the CGC, UNC, CST, IFO and SLFt along with quantitative values (FA, ADC, AD, RD) in these regions. A condensed summary of FA, ADC, AD and RD differences in statistically significant different regions of bundles between the left and right hemisphere is given in Table 8.1.

For schizophrenic patients, similar left-right differences in diffusion parameters were found as in healthy subjects (see section 8.1.3). Spatial

Table 8.2: Summary of differences in diffusion parameters for statistically significant regions of bundles from the left and right hemisphere (schizophrenic patients).

Bundle	Average quantitative values			
	FA	ADC	AD	RD
CGC	$l \gg r$	$l \approx r$	$l \gg r$	$l \ll r$
UNC	$l \ll r$	$l > r$	$l < r$	$l \gg r$
CST	$l \ll r$	$l > r$	$l \approx r$	$l \gg r$
IFO – R1	$l \ll r$	$l > r$	$l < r$	$l \gg r$
IFO – R2	$l \ll r$	$l > r$	$l \approx r$	$l \gg r$
SLFt – R1	$l < r$	$l > r$	$l > r$	$l \gg r$
SLFt – R2	$l \gg r$	$l \approx r$	$l \gg r$	$l \ll r$

location and extent of statistically significantly different regions were mostly alike to the statistically significantly different regions found in healthy volunteers. However, as spatial location and extent of these significant regions were not identical between both groups, extracted values have thus to be compared carefully, taking into account that observed differences may be due to variations in the spatial location of studied regions. In the cingulum (Figures 8.7), for example, a much smaller region was found to be statistically significant different between the left and right hemisphere compared to the region found in healthy volunteers. Due to the smaller extent of the region, the extracted, average FA, RD and ADC differences between the CGC of the left and right hemisphere were increased. For the UNC (Figures 8.8), findings between healthy volunteers and schizophrenic patients were almost identical. The location of the statistically significant regions was slightly different, which resulted in subtle changes of FA and RD. For the CST (Figures 8.9), however, no statistically significant differences between the left and right hemisphere remained after applying the FDR correction. Thus, only the statistically significant different region before applying the FDR correction was compared to the findings in healthy controls. As spatial location and extent of the observed regions differed between schizophrenic patients and healthy subjects, major changes in FA, AD and RD were found. In the IFO, two regions R1 and R2 were statistically significant different between the left and the right hemisphere (Figures 8.10). By comparing region R1 of the schizophrenic patients to the almost identical region R1 of the healthy controls, slight increases in the difference between the IFO of the right and left hemisphere were found for FA, ADC and RD. In contrast, region R2 is much larger for the schizophrenic patients, which resulted in pronounced differences of all quantitative values between schizophrenic patients and healthy volunteers. In the SLFt, again two statistically significant different regions were found (Figures 8.11). Compared to the corresponding findings in the healthy controls, both regions are located in spatially different areas of the SLFt bundle. Region R1, in particular, shares no volume with region R1 of the control group and quantitative values are thus not comparable. Region R2 is only slightly dislocated and associated with an increase of differences in all quantitative values.

On first inspection, findings for schizophrenic patients are in good agreement with the observations in healthy volunteers (see section 8.1.3). Both studies found similar differences in diffusion parameters between bundles of the left and right hemisphere. Observed differences between the two groups were thereby most likely the result of variations in spatial location and extent of these regions. However, it is noticeable that the extent of the significant regions was consistently smaller in schizophrenic patients compared to healthy volunteers, even though the significance level was adjusted to account for the smaller group size in the second study.

8.2 Assessing hemispheric white matter differences in schizophrenic patients

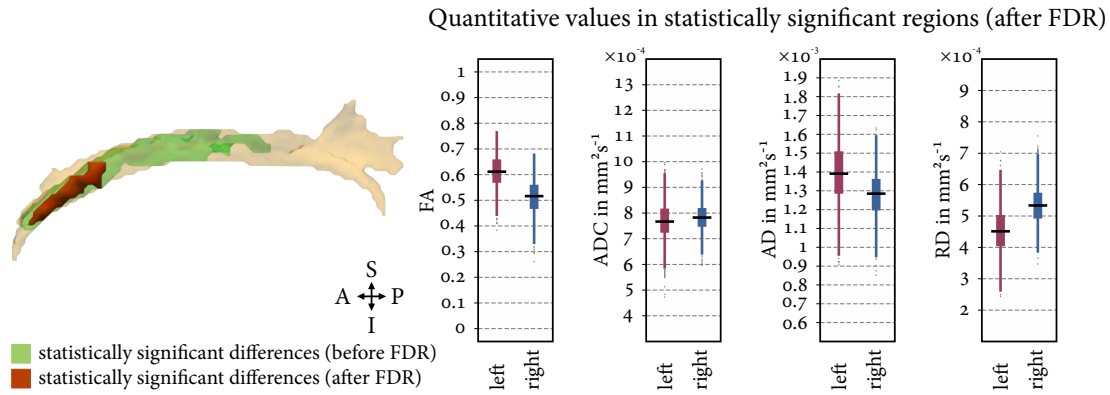


Figure 8.7: Hemispheric differences in the CGC of schizophrenic patients. The volume rendering on the left shows the analyzed CGC and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant voxels (after FDR, $p < 0.05$).

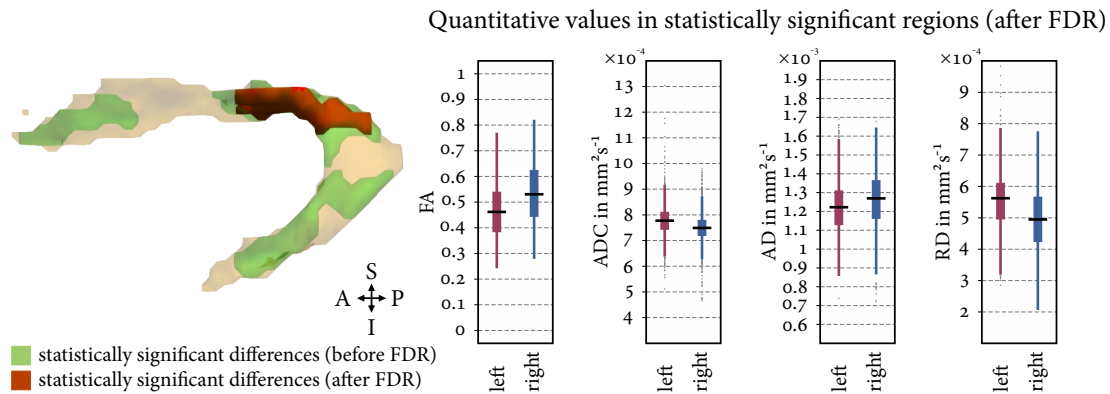


Figure 8.8: Hemispheric differences in the UNC of schizophrenic patients. The volume rendering on the left shows the analyzed UNC and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant voxels (after FDR, $p < 0.05$).

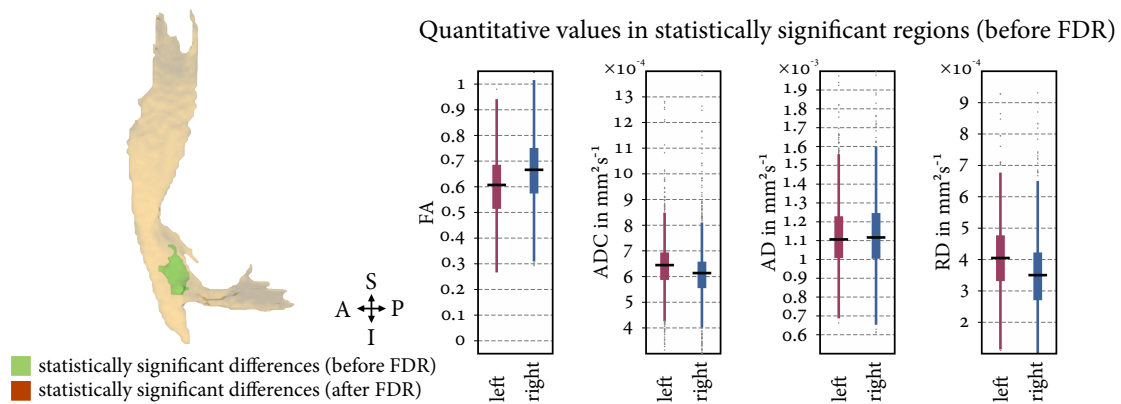


Figure 8.9: Hemispheric differences in the CST of schizophrenic patients. The volume rendering on the left shows the analyzed CST and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant voxels (before FDR, $p < 0.05$).

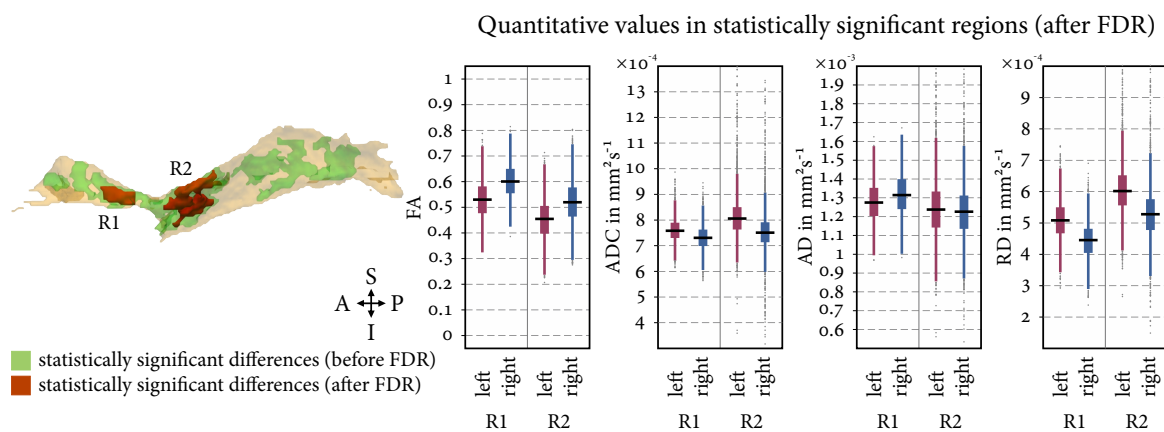


Figure 8.10: Hemispheric differences in the IFO of schizophrenic patients. The volume rendering on the left shows the analyzed IFO and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant regions R1 and R2 (after FDR, $p < 0.05$).

8.2 Assessing hemispheric white matter differences in schizophrenic patients

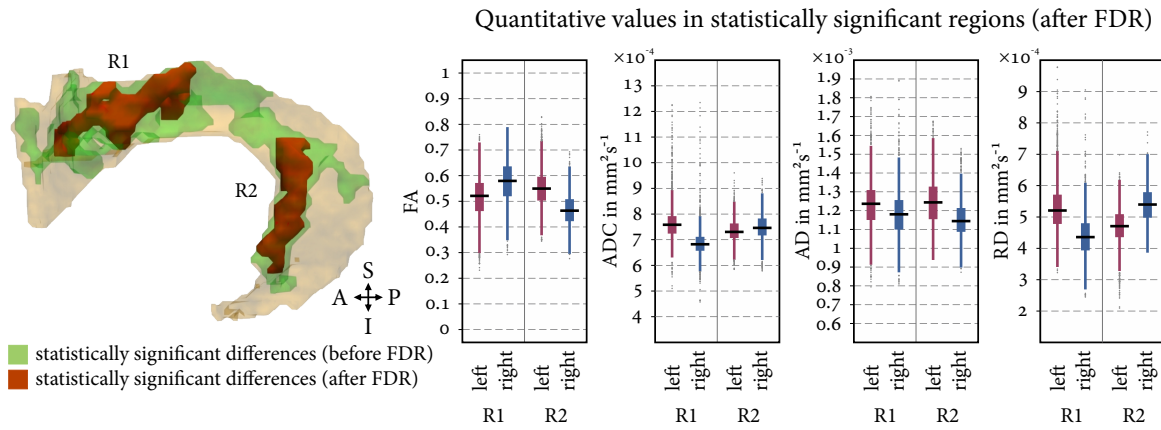


Figure 8.11: Hemispheric differences in the SLFt of schizophrenic patients. The volume rendering on the left shows the analyzed SLFt and voxels that have statistically significantly different FAs between the left and right hemisphere. Voxels with statistically significant differences are shown in green (before FDR correction) and brown (after FDR correction). The box plots on the right present quantitative values of the left and right hemisphere from all datasets in the statistically significant regions R1 and R2 (after FDR, $p < 0.05$).

Part IV

Discussion & conclusion

9

Chapter 9

Discussion and conclusion

Everything we hear is an opinion, not a fact. Everything we see is a perspective, not the truth.

— Marcus Aurelius (121–180)

Since the introduction of *diffusion weighted imaging* and the advent of fiber tractography, researchers explore novel ways to map and unravel the inherent complexity of the white matter fiber network in the human brain. While fiber tracts offer the unique opportunity to study the course and integrity of the underlying white matter microstructure in unprecedented ways, the handling and processing of the large tractography datasets is complicated and generally impaired by long processing times. The automatic parcellation of fiber tracts into bundles that represent the underlying white matter (micro-) structure is thereby of extraordinary importance for many medical applications, like delineation of tumorous tissue, radiation treatment planing or neuronavigation, but often too time consuming thus limiting the usability of fiber tractography.

In order to improve the applicability of fiber tractography, this thesis introduces novel concepts and strategies towards automatic parcellation of fiber tracts with subsequent cluster analysis. With CATSER (*cluster analysis through smartly extracted representatives*), a new method for the automated clustering of fiber tracts was presented, which exploits the intrinsic redundancy of the data to make cluster analysis applicable to large tractography datasets. By using parallel computing, random sampling and partitioning as well as novel similarity measures, cluster analysis can be performed in reasonable time. In order to improve the grouping of tracts into bundles that represent the underlying microstructure more correctly, CATSER can be used together with a white matter atlas.

Especially in the light of the recent introduction of algorithms for generating and visualizing fiber tracts with the already built-in MRI scanner software of vendors, methods and tools for fast and correct extraction of fiber bundles are becoming increasingly important to aid the medical personal in analyzing and interpreting tractography data.

Besides the application of CATSER for clustering single subject datasets in clinical routine, the presented techniques and algorithms enable fast and fully automated clustering of multiple

datasets for quantitative analyses of diffusion parameters in the white matter with *fiber bundle-driven techniques* (FDT). With these recent advances in fiber clustering, it is now feasible to extend the quantitative analysis of white matter microstructure by incorporating fiber bundles. The thesis also presented a new approach to quantitatively analyze diffusion properties in white matter fiber bundles, enabling selective analysis of individual bundles while simultaneously restricting the analysis to voxels that only belong to the fiber bundles. By projecting quantitative values onto fiber tracts prior to spatial normalization, normalization-induced interpolation between different fiber bundles are prevented.

To investigate the applicability of the presented quantitative analysis method, diffusion-associated hemispheric differences in white matter bundles of healthy volunteers and schizophrenic patients were investigated. Both studies were performed successfully and revealed spatially localized hemispheric differences in various bundles.

9.1 Cluster analysis

As the high computational complexity of many conventional clustering algorithms restricts cluster analysis to small datasets (e.g., Ding et al. (2003); Moberts et al. (2005); Maddah (2008)), this thesis introduced a new framework for the cluster analysis of large tractography datasets derived from diffusion weighted MRI data (see part II). To assess the similarity of fiber tracts, the framework contains several proximity measures that can either be used independently or in conjunction with other similarity measures to perform clustering in higher dimensions. The new clustering method CATSER allows to incorporate structural information of a white matter atlas to achieve an anatomically correct and reproducible grouping of fiber tracts. This approach combines the benefits of classification and clustering and can be considered as a hybrid technique. In addition to the identification of clusters that correspond to the classes of the atlas, CATSER also extracts additional clusters that are concealed within the dataset. If an atlas is not available (e.g., in pediatric cases (Mentzel et al., 2011)), the technique can still be used to cluster large datasets without anatomical guidance. To evaluate the developed framework (see chapter 6), 46 tractography datasets of healthy volunteers were processed and clustered with three different clustering techniques and three similarity measures (*combined distance* (CD), *Hausdorff distance* (HD) and *matched point distance* (MPD)). CATSER (with and without atlas-guidance) and conventional *hierarchical agglomerative clustering* (HAC) were used for clustering. With HAC only a small subset of 10 000 tracts was clustered due to its high computational complexity.

For the majority of the fiber bundles corresponding to classes in the atlas, consistent results were obtained as demonstrated by the assessment of the spatial agreement between clustered fiber bundles and their corresponding atlas classes. For all similarity measures, the incorporation of the volumetric atlas clearly resulted in an improved, more consistent and reproducible clustering of fiber tracts. While spatial resemblance to the atlas classes increased, variability decreased. In contrast, both other methods (CATSER without atlas and HAC) had a significantly higher spread of the values defining spatial agreement.

The comparison of the utilized similarity measures revealed consistently higher spatial agreement for HD compared to CD and MPD. Even though incorporation of shape similarity

into CD was supposed to improve clustering and increase spatial agreement between clusters and atlas classes, this assumption was not confirmed. One reason for the lower spatial agreement may be due to the fact that the shape similarity is not computed for the entire tracts but only for matched partial tract parts to determine the similarity of complete and incomplete tracts (see section 4.7). If the shape of the matched tract parts is highly similar, agreement rates are also high even if the tracts are not incomplete but belong to a different fiber bundle. Figure 6.7 shows an example where the IFO contains a substantial part that belongs to the UNC. However, in practical terms one has not only to consider spatial agreement but also computation time. Here, the best performance was achieved by using MPD for cluster analysis. While clustering with MPD was more than twice as fast compared to CD, clustering with CD was still 10 times faster than the cluster analysis with HD.

The spatial correspondence between resulting clusters and atlas classes was computed by using the number of all cluster voxels and the number of cluster voxels that intersect the atlas class (see Eq. 3.12). While this procedure takes into account voxels that are occupied by both – the cluster and the atlas class, it entirely neglects the possibility that tracts may have been falsely assigned to the fiber bundle and occupy voxels that belong to other atlas classes. To check the correctness of the obtained clusters more accurately, future studies should ideally use measures that not only consider voxels that occupy the corresponding atlas class but also voxels of the cluster that belong to different classes.

The separation of the similarity computations from the clustering permits dramatic reduction of the computation time for the cluster analysis by utilizing multiple CPUs of modern multiprocessor systems. The performance of the clustering framework was analyzed and the high speedup in a multiprocessing environment was demonstrated. The framework delivers nearly optimal speedup and performance as long as the employed similarity measures are implemented efficiently.

As an anatomically correct and consistent grouping of tracts across multiple subjects is difficult to achieve without additional anatomical information, CATSER allows incorporation of an atlas to guide the clustering process. As demonstrated in this thesis, CATSER is capable to extract bundles that are defined in the atlas and also extracts additional reasonable clusters with conventional clustering. To influence the cluster generation the spatial correspondence between tracts and atlas classes was computed and incorporated into the clustering. Since the atlas contains not only the occupied voxels for each atlas class but also additional probabilistic information, it might be reasonable to incorporate this information in the future. The atlas contains only the major parts of the bundles due to the employed ROI-based extraction technique. Even though the spatial agreement between obtained clusters and atlas classes was already high, manual generation of an atlas that captures even smallest details will further increase the spatial agreement between the clusters and the atlas bundles.

The validity and anatomical correctness of the obtained clusters is not only affected by the cluster algorithm. Both data quality and preprocessing techniques play a crucial role and influence the generation of the white matter atlas and the cluster analysis. In this study, *diffusion tensor imaging* (DTI) and deterministic tractography were only used, as such techniques are widely available and often used in clinical routine (Mentzel et al., 2011). However, the diffusion signal measured in this way is just an approximation of the underlying microstructure that in fact consists of billions of axonal connections per imaging voxel (Drachman, 2005). Even

though the diffusion tensor is a good approximation if tissues are homogeneous and axons are aligned parallel, it is insufficient in heterogeneous tissue, where multiple fiber compartments give rise to complex diffusion signals (e.g., crossing, fanning or kissing fibers). In such cases, DTI is incapable to describe the microstructure properly and tractography often fails to reconstruct meaningful fiber tracts. Tracts may be incorrect, may be disrupted along their course or may follow an unreliable trajectory that belongs to distinct nearby anatomical structures. Examples for tracking artifacts and tracts that connect wrong regions are seen in Figures 6.3 and 6.9. A bundle containing additional spurious tracts is shown in Figure 6.8. While the clustering technique can counter some of these imperfections by using outlier elimination, it will be unable to identify incorrect tracts if there is a sufficiently large number of similar, misguided tracts. If such tracts share major parts of the atlas-class volume and also have a high (partial) correspondence to the other tracts in the bundle, the atlas-guided clustering is limited in excluding these incorrect fibers.

In order to describe the diffusion in the presence of complex fiber architectures accurately more elaborate methods have to be used. Applying MRI techniques that sample the diffusion signal in many different directions with high b-values (e.g., *high angular resolution diffusion imaging* (Tuch, 2004; Descoteaux et al., 2007)), so called *orientation distribution functions* (ODFs) can be reconstructed from the data that allow to resolve regions with crossing fibers. By employing *high angular resolution diffusion imaging* in conjunction with ODF-based tractography techniques (Descoteaux et al., 2009), significant improvements in data quality and a reduction of unreasonable clusters are anticipated.

As the white matter atlas is generated on the basis of reconstructed fiber tracts, erroneous tractography will influence the atlas generation. Connections that cannot be traced reliably will be thus missing from the atlas (e.g., the connection between the uncinate fasciculus and the inferior frontal gyrus in Figure 6.8). Considering heterogeneous tissue where multiple fiber compartments exist (e.g., crossing fibers), a voxel-based probabilistic atlas is limited as probabilities will be significantly reduced in regions with crossing fibers. An orientation-dependent atlas that takes into account the fiber orientation in each voxel by using, for example, ODFs, will be the next reasonable step to further enhance the clustering and to prevent the bundling of misguided tracts. In addition, a grey matter atlas might also be employed to improve the clustering. In this case, a weighting factor might be based on the distance between the tracts and their closest grey matter region.

The atlas-guided clustering approach aims at fast and consistent extraction of fiber bundles from multiple subjects by employing a white matter atlas. Compared to recently presented techniques (e.g. Visser et al. (2011); Guevara et al. (2011); Garyfallidis et al. (2012)), it enables anatomically correct extraction of bundles that are defined in the atlas as well as extraction of additional bundles that are not available in the atlas. While Guevara et al. (2011) uses a reasonably fast approach that is not based on clustering of tracts but on clustering of voxels, Visser's algorithm employs partitioning and repeated hierarchical clustering to achieve a grouping of tracts. Even though Visser's technique (Visser et al., 2011) works quite well for the clustering of fiber tracts, it is limited by the long computation times, which makes it impractical to cluster large numbers of datasets. However, although such similarity-based approaches are useful for grouping fiber tracts, good correspondence of the clustered bundles to the true anatomical bundles cannot be guaranteed due to the lack of anatomical information. As a

result, Guevara et al. (2012) extended their voxel clustering approach to increase anatomical correspondence by incorporating a white matter atlas. First, agglomerative voxel clustering is performed until many thousand, clustered voxels remain. This multitude of clusters is then classified using a manually labeled white matter atlas, derived from a set of clustered datasets. A different approach for an anatomically correct extraction of white matter fiber bundles was proposed by Yendiki et al. (2011) with the TRACULA framework. Instead of using cluster analysis to retrospectively bundle fiber tracts from whole brain tractography datasets, TRACULA aims to reconstruct bundles separately during the fiber tracking itself. Hereby, probabilistic tractography is used to reconstruct only the tracts of interest by restricting the tracking using anatomically a-priori-defined *regions of interest* as defined by Wakana et al. (2007).

Besides the employed algorithms, cluster analysis also depends fundamentally on the employed similarity measures that define the criteria to distinguish fiber tracts. Popular measures such as the *Hausdorff distance* (Moberts et al., 2005), the *Chamfer distance* (Corouge et al., 2004) or shape-based similarity measures (Batchelor et al., 2006) are often used, but neglect partially overlapping (Wassermann et al., 2010) or incomplete fibers. To take into account partially overlapping tracts, Wassermann et al. (2010) proposed a mathematical framework that modeled tracts and fiber bundles as Gaussian processes in voxel space. In Wassermanns framework cluster are successively formed and the Gaussian processes of tracts are merged in each iteration. This kind of processing has some resemblance to the presented atlas-guided clustering.

Finally, the scalability of the clustering framework was analyzed with respect to the number of employed CPU cores. The results revealed that the computationally most demanding parts are the computations of fiber tract similarities, while clustering with CATSER itself is quite fast. An in-deep analysis demonstrated that the clustering framework enables fast processing with minimal computational overhead. Even when all CPU cores of a system are employed, the achieved speedup is nearly optimal. However, the results also indicate that speedup of the framework depends on the utilized similarity measure. This was confirmed by additional experiments, which revealed that a reduced speedup for certain similarity measures is primarily related to the computer system's memory management. It can only be assumed that the degraded multiprocessing efficiency is the result of a memory bandwidth bottleneck because of the system's inability to handle excessive memory (de-) allocations that may occur during the computation of certain similarity measures. To minimize such performance degradation, memory (re-) allocation during the computation of similarities has to be minimized. Although the presented framework already offers good performance, an additional gain may be achieved by developing more sophisticated algorithms for determining fiber tract similarity. To investigate the prospects of alternative hardware architectures, the computation of the *Hausdorff distance* was implemented for *graphics processing units* (GPUs) that are based on a stream processing architecture (Owens et al., 2008; NVIDIA, 2010). By using different GPU systems for the computation of the similarities, the achievable performance was assessed and compared to processing with conventional multi-core CPU systems. Even though GPU systems demonstrated their potential to reduce the processing time notably, they are not suited for complex similarity measure that contain conditional statements. Due to the conditional statements, branching and stalling of processing cores is

inevitable and degrades processing performance. Besides the computation of similarities on GPU systems, it might also be worthwhile to develop and investigate GPU-aided clustering methods. They may be used to replace the clustering of the random sample (see Figure 3.2, steps 3–7) with an alternative, more efficient clustering method. In the future, this may ultimately help to achieve some kind of real-time clustering.

9.2 Quantitative analysis of fiber bundles

Alterations in diffusivity are often indicators for structural changes that are related to either natural or pathological causes. To assess altered diffusion in the white matter microstructure, a novel method for quantitative group-based analyses was introduced in this thesis (see part III). With the presented approach, diffusion properties (e.g., FA, ADC) are not studied in the whole brain, but across individual *bundles of interest* (BOI). The clustering framework is thereby used to automatically partition the fiber tracts of multiple datasets into bundles that correspond to known anatomic white matter structures (see Part II). In order to prevent normalization-induced interpolation of quantitative values of the diffusion parameters at the boundaries of tissue structures, these values are projected onto the tracts before spatial normalization¹. As fiber tracts are defined by real-valued points in 3D space, the spatial normalization is only a coordinate transformation that changes the course of the tracts but preserves the values that are attached to the tract points. For the individual voxel-wise analysis of each *bundle of interest*, the attached values are gridded separately for each bundle before the subsequent statistical analysis, thereby avoiding interpolation of values from different, adjacent bundles. During the individual analysis of each *bundle of interest*, only corresponding fiber bundles from all datasets of the group are used. This suppresses contributions of voxels that do not belong to the analyzed white matter structures, thereby minimizing the negative influence of other structure on the statistical voxel-wise analysis.

With respect to other techniques for quantitative analysis (see section 2.5), the presented method can be considered a hybrid *fiber bundle-driven technique* (FDT). It is a comprehensive approach that aims to improve conventional analyses techniques (VBM (Ashburner and Friston, 2000), TBSS (Smith et al., 2006)) by incorporating additional information about subject-specific white matter anatomy. Like VBM, the presented technique performs a voxel-wise analysis, but uses subject-specific fiber bundles to delineate regions for the statistical analysis. With VBM, different tissue types and anatomical structures are not considered adequately, which often results in statistical errors that are foremost notably at tissue boundaries. With the presented FDT approach, however, each BOI is processed separately and only anatomically corresponding fiber bundles from all datasets are used.

The complications with VBM culminated in the development of TBSS, which performs the statistical analysis along an average FA skeleton. The statistical analysis of TBSS is notably more robust compared to VBM, but is also restricted to the white matter skeleton, which represents only the most prominent area of the fiber bundles, extracted from the averaged FA maps of all datasets. Alterations that occur outside the FA skeleton or in the periphery

¹The spatial normalization transfers the individual datasets to the common template space.

of a white matter structure may remain undetected. With the presented *fiber bundle-driven technique*, peripheral regions can still be investigated whilst any negative influence of other, unrelated structures on the analysis is prevented. The introduced FDT approach can essentially be considered a middle ground between VBM and TBSS that combines the benefits of both methods.

Compared to other *fiber bundle-driven techniques* that can only be used to study diffusion along the course of the BOI's mean tract (see sections 2.5 and 7.1 for an overview), the presented FDT technique has the merit that the analysis of the BOI is not restricted to tubular structures but can be performed with fiber bundles of arbitrary shapes.

While the presented technique is an initial step to enhance quantitative analyses by incorporating complementary anatomical information, limitations and challenges still exist. As fiber bundles are utilized as additional anatomical information, the analysis is restricted to white matter structures and processing is more complex compared to conventional techniques (e.g., VBM, TBSS). For the presented technique, tractography is necessary to reconstruct the white matter fiber network, before the obtained fiber tracts are grouped into bundles. The analysis is thereby performed on the basis of fiber bundles and thus relies on the correctness of extracted tracts, which is in return governed by the utilized tractography algorithm. If the course of the tracts does not represent the underlying white matter structure, potential tracking artifacts (e.g., tracts that connect unrelated regions) may introduce additional interpolation of quantitative values and may affect the subsequent analysis. By using advanced acquisition, reconstruction and tracking methods, the quality of the reconstructed fiber network will improve and ultimately benefit the analysis (see the previous section 9.1 for a more detailed discussion on advanced reconstruction and tracking methods).

Besides fiber tracking, the subsequent grouping of tracts is essential for the statistical analysis as it defines the bundles that are later analyzed. If the grouping into bundles fails, the quantitative analysis cannot be performed. If extracted bundles are incorrect and/or contain additional tracts, the quantitative analysis can usually still be performed but will be impaired by interpolation effects from adjacent structures. While the grouping of the fiber bundles can be performed with manual and (semi-) automatic techniques, automatic methods are usually faster and less prone to errors due to the minimized user interactions (e.g., automatic fiber clustering, see part II). In some cases however, when bundles only consist of few tracts, the automatic extraction of bundles may fail (e.g., the CGH, see section 6.5.1). In such cases, manual methods may be applied instead to still obtain bundles.

Like all automatic methods for quantitative analysis, the presented FDT approach relies on spatial normalization to superimpose corresponding brain structures of different datasets and to account for inter-subject variability of brain anatomy. For the analysis it is essential that the investigated bundles are properly aligned and corresponding regions are actually compared. If certain brain structures are not superimposed adequately, the quantitative analysis might be problematic in these bundles. In order to cope with such cases, it might be feasible to use bundle-wise coregistration in which the normalization is performed separately for corresponding bundles of the different datasets. If bundle-wise normalization is performed in addition to the conventional spatial normalization, small misalignments may be compensated even if the spatial normalization worked well and already achieved good alignment. However, to utilize a bundle-wise normalization, bundles have to be extracted in advance and the region that is

occupied by the bundles has to be delineated in the 3D volumes of the datasets. If the bundle delineation is insufficient this approach will most likely also fail and result in an inadequate superimposition of bundles. Additional normalization errors may also be introduced as the spatial normalization cannot use anatomical landmarks for guidance during bundle-wise normalization.

In order to avoid interpolation effects due to the spatial normalization, quantitative values of diffusion parameters are projected onto the fiber tracts prior to the normalization. For the quantitative analysis, tracts are clustered, obtained fiber bundles are rasterized and the values, still attached to the bundles, are gridded to a regular grid for each fiber bundle individually (see section 7.2.1). In regions where multiple fiber compartments exist and fiber bundles overlap, the bundle-wise gridding prevents the interpolation of values from overlapping bundles (see section 7.4.2, first experiment). By investigating gridded FA values that were attached before (FA_{pre}) and after spatial normalization (FA_{post}), the occurrence of normalization-induced interpolation effects was demonstrated by showing differences in the distributions of analyzed values (cmp. section 7.4.2, second experiment). These effects were thereby only shown indirectly by comparing differences between conventional gridding of the whole dataset and bundle-wise gridding of individual bundles for FA_{pre} and FA_{post} . The obtained voxel value differences between conventional and bundle-wise gridding were larger for FA_{pre} than for FA_{post} , which is a strong indicator that these differences are truly the result of interpolation due to the spatial normalization (see section 7.4.2).

Additional gridding-related differences were also observed for conventional and bundle-wise gridding (see section 7.4.2, second experiment), which are the result of variations in the tracts that were gridded with conventional and bundle-wise gridding. In voxels that are occupied by tracts of multiple bundles, the number of tracts that contribute to the voxels during bundle-wise gridding is reduced compared to the conventional gridding that takes into account the different contributions of all tracts, irrespective to which bundle the tracts belong to. Considering both interpolation effects, i.e., normalization-induced and gridding-related interpolation of quantitative values, the question remains to what extent these effects truly influence the subsequent statistical analysis. To investigate the impact of the spatial normalization and the gridding on the analysis, additional experiments have to be performed in future studies.

One further issue is related to the gridding resolution and the chosen tract-related parameters (i.e., number of tracts, distance between tract points, see sections 7.3.1 and 7.4.1). In order to preserve quantitative values, the number of tract tracts and the distance between tract points has to be chosen adequately depending on the grid resolution. With respect to the tract point distance, one has to consider that the analysis is not performed in native space but in the atlas space. As the normalization results in a coordinate transformation that changes the course of the tracts and thus the distance between tract points, it might be necessary to reduce the tract point distance to avoid misrepresentation of quantitative values in the atlas space.

As fiber bundles are instrumented for the presented quantitative analysis technique, it is inevitable to discuss the resolution that is used for the voxel-wise statistical analysis. Tracts of the fiber bundles are defined by real-valued points in 3D space, which makes it mandatory to rasterize the tracts and grid the attached values to a regular grid (step 5 in Figure 7.3) prior to

the statistical analysis (step 7 in Figure 7.3), The rasterization of the tracts can be performed for arbitrary resolutions and implicitly defines the resolution for which the subsequent statistical analysis is performed. However, which gridding resolution is optimal for the analysis is debatable. While an analysis with a resolution that is lower than the resolution in the native space of the datasets is usually not desired², a higher resolution may help to improve the delineation of regions with statistically significant differences (see section 7.4.2). In practice, however, very high resolutions (e.g., 0.4 mm³ isotropic resolution) are hardly suitable, as higher resolutions requires more tracts as well as shorter tract point distances to prevent a misrepresentation of quantitative values after gridding (see section 7.4.1). As the projection of values onto the tracts is performed in native space, it is unlikely that a substantially higher gridding resolution will aid the statistical analysis in any other way than to improve the delineation of fiber bundles. However, if *super resolution techniques* are employed (e.g., *tract density imaging* (Calamante et al., 2011)) or other contrasts (e.g., T₁, T₂) are used that have a higher resolution than the conventional DTI on which the fiber tracking is based on, the statistical analysis should also be performed with a high resolution to preserve to original, high resolution.

To lower the overall processing time, it might be feasible to reduce the number of tracts in the datasets. However, considering the dependency on the parameters (see section 7.4.1) and the previously observed gridding-related interpolation effects (see section 7.4.2 and above), tracts should not be removed randomly as this might culminate in additional gridding errors if tracts are removed from regions where few tracts exist. Instead, tracts should be removed if they traverse through regions that contain plenty of tracts. However, if such a density-based tract removal scheme is applied, subsequent analysis that aim to investigate density differences will be rendered invalid and will most certainly fail to detect differences. Consider for example the study by O'Donnell et al. (2009) who found differences in the number of the tracts between the CGC of the left and right hemisphere. In such a case, it is most likely that a density-based removal of tracts will remove tracts from bundles that consist of many tracts, thereby removing differences that would have been observed otherwise.

With the presented method developed in this thesis, the statistical analysis is performed separately for each fiber bundle to detect statistically significant differences in diffusion properties between two or more groups of subjects (e.g., patients vs. healthy controls). During the analysis, adequate statistical hypothesis tests have to be performed in all voxels that are occupied by the *bundle of interest* to determine if encountered differences between both groups are statistically significant. To guarantee that results are valid, one has to consider that the number of datasets that contribute to each voxel fluctuates with its location in the bundle (see section 7.2.5). In some voxels, the requirements for a certain statistical hypothesis test might not be satisfied, rendering the tests in these voxels invalid. Instead of classical tests that are only based on assumptions about the probability distributions (e.g., *Students t-test*, *z-test*), exact hypothesis tests may be more appropriate to be used for the statistical analysis (e.g., *permutation tests*).

²Though it may be considered some kind of smoothing due to the occurring interpolation.

9.3 Applications

The applicability of the presented *fiber bundle-driven technique* for quantitative analysis was tested by conducting two studies that investigated diffusion-associated hemispheric differences in white matter bundles of 46 healthy volunteers (study 1) and 15 schizophrenic patients (study 2, see chapter 8). For both studies, the processing was almost identical. To account for the fewer subjects in study 2, the significance level was adjusted to $p < 0.05$. Though both initial studies only explored diffusion differences between cerebral hemispheres, the technique can also be used to detect alterations of diffusion in white matter bundles of both hemispheres between groups of subjects or subgroups of patients.

In both studies, noticeable differences in diffusion were found between corresponding bundles of the left and right hemisphere. After correcting for multiple comparisons with the *false discovery rate* – FDR (Benjamini et al., 2006), results of both studies were still statistically significantly different in regions of the gyrus part of the cingulum (CGC), uncinate fasciculus (UNC), inferior fronto-occipital fasciculus (IFO) and the temporal part of the superior longitudinal fasciculus (SLFt). In healthy volunteers, statistically significant differences were also observed in regions of the corticospinal tract (CST). In the CST of schizophrenic patients, however, differences were not significant after correcting for multiple comparisons. Nevertheless, the result indicates at least a trend of differences in quantitative values between the CST of the left and right hemisphere.

Overall, both studies found similar differences in the diffusion parameters between bundles of the left and right hemisphere. Differences between the two studies were primarily related to variations in the spatial location and the extent of certain statistically significantly different regions (see sections 8.1.3 and 8.2.3). Quantitative values that were extracted from these regions were almost identical for healthy volunteers and schizophrenic patients. Even though the significance level was set to $p < 0.05$ for study 2, the extent of the significant regions was consistently smaller in schizophrenic patients compared to healthy volunteers. This might be an indication that hemispheric asymmetries are less pronounced in schizophrenic patients, which is supported by a study of Park et al. (2004), who also investigated hemispheric diffusion differences between healthy volunteers ($N = 32$) and schizophrenic patients ($N = 23$) and found similar results. Nevertheless, additional investigations (e.g., with larger group sizes) are certainly necessary to draw any final conclusions whether these observed differences are specific for schizophrenia. The large difference in the number of healthy volunteers ($N = 46$) and schizophrenic patients ($N = 15$) may also explain why the extent of statistically significant different regions was consistently smaller for schizophrenic patients.

Overall, detected differences and non-existing differences between healthy volunteers and schizophrenic patients are in line with the literature. However, as observed by Shenton et al. (2001) and Kubicki et al. (2007), in their review articles, results for schizophrenia are heterogeneous and for every study that found differences between healthy volunteers and schizophrenic patients another study exist that found none.

At least for healthy volunteers, differences in diffusion parameters between bundles of the left and right hemisphere are obviously not the product of pathological processes but occur naturally due to differences at a (micro-) cellular level. Considering the hemispheric asymmetry and bilateralization in both structure and function that have been reported over

the last 150 years (see section 8.1.1), it is only reasonable that such profound differences manifest itself in altered diffusivity. Even though Miller (1996) proposed a hypothesis that related functional differences to alterations in size and structure of axons, a study in rats by Partadiredja et al. (2003) found no axonal differences between the left and right hemisphere that supports this hypothesis. However, the study was only performed in rats which might explain the negative results. In contrast, Highley et al. (2002) performed a *post mortem* study and investigated the uncinate fasciculus (UNC) in healthy controls and schizophrenic patients. They observed no significant differences in the UNC between schizophrenic patient and healthy controls, but found, in both groups, hemispheric asymmetries and a larger UNC in the right hemisphere that consisted of 33 % more fibers, indicating a potentially greater right-sided connectivity of the UNC. Their observations are overall in line with the results for the UNC in both studies of this thesis that detected higher *fractional anisotropy* (FA) and *axial diffusivity* (AD), but also lower *radial diffusivity* (RD) (see above and sections 8.1.3, 8.2.3). Here, the observed lower RD in the right hemisphere may indicate higher packing density of axons in the UNC, while the higher AF and AD support the hypothesis of greater connectivity in the right UNC. Histological studies of hemispheric differences are unfortunately rare for other bundles. Nevertheless, the observed differences in diffusion parameters between bundles of the left and right hemisphere (both studies) as well as the identified patterns of hemispheric asymmetry are in accordance with the literature (Fillard et al. (2003); Park et al. (2004); Gong et al. (2005); O'Donnell et al. (2009) and Thiebaut de Schotten et al. (2011)). The SLFt findings, for example, are supported by two studies. While Eluvathingal et al. (2007) and Catani et al. (2007) observed lower left than right FA in the anterior segment of the SLFt, Catani et al. (2007) also found higher left than right FA in its long segment (see Figure 8.6). By considering the locations of the statistically significantly different regions (R₁ = anterior segment and R₂ = long segment), these findings are in good agreement with the observations in this thesis (see sections 8.1.3 and 8.2.3).

Interpretation of quantitative diffusion tensor-based values (e.g., FA, ADC, AD, RD), however, is not free of pitfalls and has to be performed cautiously (Wheeler-Kingshott and Cercignani, 2009). While these parameters characterize important properties of the diffusion tensor, they can only be related to underlying biophysical properties (e.g., fiber density, axonal myelination) if the diffusion tensor model is valid. In regions of multiple fiber compartments (e.g., crossing fibers), the assumption of Gaussian diffusion is no longer valid and the diffusion tensor model breaks down. To account for this problem, the different fiber configurations have to be separated and analyzed individually (see the method by Caan et al. (2010)) or more advanced techniques, such as *high angular resolution diffusion imaging* (see section 9.1) have to be used to probe the diffusion microstructure and resolve complex fiber configurations. Further approaches that use a *cone of uncertainty* to characterize directional uncertainty of eigenvectors are also well suited (Koay et al., 2008; Jeong and Anderson, 2008; Wheeler-Kingshott and Cercignani, 2009). In the studies conducted in this thesis (see above), the investigated fiber bundles traversed, for the most part, regions with one principal direction of diffusion and no fiber crossings. Even though the pitfalls stated before have to be considered carefully, they are only of minor importance for the presented studies.

9.4 Conclusion

The present thesis introduced novel concepts and promising strategies for the cluster analysis of large tractography datasets and the quantitative analysis of diffusion parameters in multi-subject imaging studies. With CATSER (*cluster analysis through smartly extracted representatives*), a new method was presented that enables fully automated grouping of fiber tracts into bundles. The approach uses random sampling of tracts to exploit the intrinsic data redundancy and partitioning to separate the data into manageable pieces. The sensitivity to outliers is reduced by incorporating *local outlier factors* that provide insight into the structural organization of the data. Moreover, multithreaded parallel programming is applied to improve cluster analysis performance and to achieve grouping of tracts in a short time. By including a white matter atlas, the grouping of tracts is enhanced and bundles are obtained that represent the underlying microstructure more correctly. The functioning of the introduced cluster analysis framework was investigated and its performance in a multiprocessing environment was demonstrated.

As fast and consistent extraction of fiber bundles supports multiple medical applications, its benefit for quantitative analyses of diffusion properties was demonstrated by introducing fiber bundles as additional anatomical information into a new *fiber bundle-driven technique*. With this new approach, parameters of diffusion are analyzed quantitatively in individual bundles while simultaneously restricting the analysis to voxels only belonging to the *fiber bundles of interest*. This incorporation of subject-specific information about white matter structures facilitates the minimization of interpolation effects and furthermore avoids the adverse influence of different, adjacent structures on the statistical analysis. It ultimately enables the analysis of diffusion properties in an unprecedented manner. The potential of this novel technique was demonstrated with two experimental studies that investigated hemispheric differences of diffusion-related parameters in healthy volunteers and schizophrenic patients. Extending the presented approach to enable fiber bundle-based quantification of other biomarkers, such as T_1 , T_2 , R_2 , R_2^* or *quantitative susceptibility mapping* (Schweser et al., 2011, 2012), will be a promising, complementary approach to gain insight into diverse disease states of the human brain that are associated with white matter alterations.

With the introduction of these new techniques, further steps were taken to make cluster analysis feasible for a variety of applications and to enhance quantitative analysis by incorporating additional information about white matter microstructure. As the introduction of these techniques opened new possibilities to investigate white matter diseases and characterize abnormal and pathological conditions, further studies are already being planned.

Bibliography

- Abdi H (2007) The Bonferroni and Šidák Corrections for Multiple Comparisons. In: *Encyclopedia of measurement and statistics*. Thousand Oaks (CA):Sage, 103–107.
- Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD (1994) From Single Cells to Multicellular Organisms. In: *Molecular Biology of the Cell*. Garland Publishing Inc., 3rd edition.
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P (2008) *Molecular Biology of the Cell*. Garland Science, 5th edition.
- Alexander AL (2011) Deterministic White Matter Tractography. In: *Diffusion MRI: Theory, Methods and Applications*. Oxford Univ. Press, New York ; Oxford, 383–395.
- American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR)*. American Psychiatric Association.
- Andreopoulos B, An A, Wang X and Schroeder M (2009) A roadmap of clustering algorithms: finding a match for a biomedical application. *Brief Bioinform* 10(3):297–314.
- Ardekani BA, Tabesh A, Sevy S, Robinson DG, Bilder RM and Szeszko PR (2011) Diffusion tensor imaging reliably differentiates patients with schizophrenia from healthy volunteers. *Hum Brain Mapp* 32(1):1–9.
- Ashburner J and Friston KJ (2000) Voxel-based morphometry – the methods. *Neuroimage* 11(6):805–821.
- Ashburner J and Friston KJ (2001) Why voxel-based morphometry should be used. *Neuroimage* 14(6):1238–1243.
- Assaf Y and Basser PJ (2005) Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *Neuroimage* 27(1):48–58.
- Assaf Y, Blumenfeld-Katzir T, Yovel Y and Basser PJ (2008) AxCaliber: a method for measuring axon diameter distribution from diffusion MRI. *Magn Reson Med* 59(6):1347–1354.
- Attalla E and Siy P (2005) Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. *Pattern Recogn* 38(12):2229–2241.
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A and Gee JC (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54(3):2033–2044.

Bibliography

- Bahn MM (1999) Comparison of scalar measures used in magnetic resonance diffusion tensor imaging. *J Magn Reson* 139(1):1–7.
- Basser PJ (1995) Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR Biomed* 8(7-8):333–344.
- Basser PJ, Mattiello J and Le Bihan D (1994) MR diffusion tensor spectroscopy and imaging. *Biophys J* 66(1):259–267.
- Basser PJ, Pajevic S, Pierpaoli C, Duda J and Aldroubi A (2000) In vivo fiber tractography using DT-MRI data. *Magn Reson Med* 44(4):625–632.
- Basser PJ and Özarslan E (2011) Anisotropic Diffusion: From the Apparent Diffusion Coefficient to the Apparent Diffusion Tensor. In: *Diffusion MRI: Theory, Methods and Applications*. Oxford Univ. Press, New York ; Oxford, 79–91.
- Batchelor PG, Calamante F, Tournier JD, Atkinson D, Hill DLG and Connelly A (2006) Quantification of the shape of fiber tracts. *Magn Reson Med* 55(4):894–903.
- Beaulieu C (2002) The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed* 15(7-8):435–455.
- Behrens TEJ, Berg HJ, Jbabdi S, Rushworth MFS and Woolrich MW (2007) Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage* 34(1):144–155.
- Bélair EL, Vallée J and Robitaille R (2010) In vivo long-term synaptic plasticity of glial cells. *J Physiol* 588(7):1039–1056.
- Benjamini Y and Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Statist Soc B* 57(1):289–300.
- Benjamini Y, Krieger AM and Yekutieli D (2006) Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3):491–507.
- Bezdek JC (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA.
- Blashfield RK and Aldenderfer MS (1978) The literature on cluster analysis. *Multivar Behav Res* 13(3):271–295.
- Bloch F (1946) Nuclear Induction. *Phys Rev* 70(7-8):460–474.
- Bookstein FL (2001) "Voxel-based morphometry" should not be used with imperfectly registered images. *Neuroimage* 14(6):1454–1462.
- Brenner S (1974) The genetics of *Caenorhabditis elegans*. *Genetics* 77(1):71–94.

- Breunig MM, Kriegel HP, Ng RT and Sander J (2000) LOF: Identifying Density-Based Local Outliers. In: Chen W, Naughton JF and Bernstein PA, editors, *SIGMOD Conference*, ACM, 93–104.
- Broca P (1865) Sur le siège de la faculté du langage articulé. *Bull Soc Anthropol Paris* 6(1):377–393.
- Brown R (1828) A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Phil Mag* 4:161–173.
- Brun A, Knutsson H, Park HJ, Shenton ME and Westin CF (2004) Clustering Fiber Traces Using Normalized Cuts. *Lect Notes Comput Sci* 3216/2004(3216):368–375.
- Bullmore ET, Frangou S and Murray RM (1997) The dysplastic net hypothesis: an integration of developmental and dysconnectivity theories of schizophrenia. *Schizophr Res* 28(2-3):143–156.
- Caan MWA, Khedoe HG, Poot DHJ, den Dekker AJ, Olabarriaga SD, Grimbergen KA, van Vliet LJ and Vos FM (2010) Estimation of diffusion properties in crossing fiber bundles. *IEEE Trans Med Imaging* 29(8):1504–1515.
- Calamante F, Tournier JD, Heidemann RM, Anwender A, Jackson GD and Connelly A (2011) Track density imaging (TDI): validation of super resolution property. *Neuroimage* 56(3):1259–1266.
- Callaghan PT (1991) *Principles of Nuclear Magnetic Resonance Microscopy*, volume 19. Oxford University Press.
- Callaghan PT (2011) Physics of Diffusion. In: *Diffusion MRI: Theory, Methods and Applications*. Oxford Univ. Press, New York ; Oxford, 45–65.
- Carpenter GA and Grossberg S (1988) The ART of adaptive pattern recognition by a self-organizing neural network. *Computer* 21:77–88.
- Catani M, Allin MPG, Husain M, Pugliese L, Mesulam MM, Murray RM and Jones DK (2007) Symmetries in human brain language pathways correlate with verbal recall. *Proc Natl Acad Sci U S A* 104(43):17163–17168.
- Catani M and Mesulam M (2008) The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex* 44(8):953–961.
- Catani M and Thiebaut de Schotten M (2012) *Atlas of Human Brain Connections*. Oxford University Press, 519 pages.
- Chao TC, Chou MC, Yang P, Chung HW and Wu MT (2009) Effects of interpolation methods in spatial normalization of diffusion tensor imaging data on group comparison of fractional anisotropy. *Magn Reson Imaging* 27(5):681–690.

Bibliography

- Clark CA and Byrnes T (2011) DTI and Tractography in Neurosurgical Planning. In: *Diffusion MRI: Theory, Methods and Applications*. Oxford Univ. Press, New York ; Oxford, 588–607.
- Clarke J (2012) Neurons in the brain. Wellcome Images, accessed on the 26th of August 2012.
URL: www.wellcome.ac.uk/Education-resources/Teaching-and-education/Big-Picture/All-issues/The-Cell/Image-galleries-Aspects-of-imaging/WTDV030842.htm
- Cohen-Tannoudji C, Diu B and Laloe F (2006) *Quantum Mechanics*. Wiley-Interscience, 1524 pages.
- Conturo TE, Lori NF, Cull TS, Akbudak E, Snyder AZ, Shimony JS, McKinstry RC, Burton H and Raichle ME (1999) Tracking neuronal fiber pathways in the living human brain. *Proc Natl Acad Sci USA* 96(18):10422–10427.
- Corballis MC (2009) The evolution and genetics of cerebral asymmetry. *Philos Trans R Soc Lond B Biol Sci* 364(1519):867–879.
- Corouge I, Gouttard S and Gerig G (2004) Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In: *IEEE Int Symp Biomed Imag*, IEEE, 344–347.
- Crank J (1975) *The mathematics of diffusion*. Clarendon Press, Oxford University Press, 2nd edition, 424 pages.
- Dauguet J, Peled S, Berezovskii V, Delzescaux T, Warfield SK, Born R and Westin CF (2007) Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain. *Neuroimage* 37(2):530–538.
- Descoteaux M, Angelino E, Fitzgibbons S and Deriche R (2007) Regularized, fast, and robust analytical Q-ball imaging. *Magn Reson Med* 58(3):497–510.
- Descoteaux M, Deriche R, Knösche TR and Anwander A (2009) Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE Trans Med Imaging* 28(2):269–286.
- Ding Z, Gore JC and Anderson AW (2003) Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *Magn Reson Med* 49(4):716–721.
- Drachman DA (2005) Do we have brain to spare? *Neurology* 64(12):2004–2005.
- Durstenfeld R (1964) Algorithm 235: Random permutation. *Commun ACM* 7(7):420.
- Einstein A (1905) Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann Phys* 322(8):549–560.
- Eluvathingal TJ, Hasan KM, Kramer L, Fletcher JM and Ewing-Cobbs L (2007) Quantitative diffusion tensor tractography of association and projection fibers in normally developing children and adolescents. *Cereb Cortex* 17(12):2760–2768.

- Everitt BS, Landau S, Leese M and Stahl D (2011) *Cluster Analysis*. Wiley Series in Probability and Statistics.
- Fick A (1855) Ueber Diffusion. *Ann Phys* 170(1):59–86.
- Fillard P, Gilmore J, Piven J, Lin W and Gerig G (2003) Quantitative analysis of white matter fiber properties along geodesic paths. *Med Image Comput Comput Assist Interv* :16–23.
- Filley C (2001) *The Behavioral Neurology of White Matter*. Oxford University Press.
- Findlen P (1998) A history of the brain. published online, accessed on the 28th of August 2012.
URL: <http://www.stanford.edu/class/history13/earlysciencelab/body/brainpages/brain.html>
- Fisher RA and Yates F (1948) *Statistical tables for biological, agricultural and medical research*. Oliver & Boyd, 3rd edition.
- Flynn M (1972) Some computer organizations and their effectiveness. *IEEE Trans Comput* 100(9):948–960.
- Forgy E (1965) Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification. *Biometrics* 21(3):768–769.
- Friston KJ (1998) The disconnection hypothesis. *Schizophr Res* 30(2):115–125.
- Friston KJ (2002) Dysfunctional connectivity in schizophrenia. *World Psychiatry* 1(2):66–71.
- Fuxe K, Borroto-Escuela DO, Romero-Fernandez W, Diaz-Cabiale Z, Rivera A, Ferraro L, Tanganelli S, Tarakanov AO, Garriga P, Narváez JA, Ciruela F, Guescini M and Agnati LF (2012) Extrasynaptic neurotransmission in the modulation of brain function. Focus on the striatal neuronal-glia networks. *Front Physiol* 3:136.
- Galaburda AM, LeMay M, Kemper TL and Geschwind N (1978) Right-left asymmetries in the brain. *Science* 199(4331):852–856.
- Garyfallidis E, Brett M, Correia MM, Williams GB and Nimmo-Smith I (2012) QuickBundles, a Method for Tractography Simplification. *Front Neurosci* 6:175.
- Gazzaniga MS (2005) Forty-five years of split-brain research and still going strong. *Nat Rev Neurosci* 6(8):653–659.
- Geschwind N (1979) Specializations of the human brain. *Sci Am* 241(3):180–199.
- Geschwind N and Galaburda A (1987) *Cerebral lateralization, biological mechanisms, associations, and pathology*. Cambridge, MA: Bradford Books.
- Ghosh P (2009) *Colloid and Interface Science*. PHI Learning Private Limited.
- Giorgio A, Watkins KE, Douaud G, James AC, James S, Stefano ND, Matthews PM, Smith SM and Johansen-Berg H (2008) Changes in white matter microstructure during adolescence. *Neuroimage* 39(1):52–61.

Bibliography

- Goldsmith AD (2011) Neuronal Laterality in *Caenorhabditis elegans*: Morphological and Functional Aspects. Ph.D. thesis, Columbia University.
- Gong G, Jiang T, Zhu C, Zang Y, Wang F, Xie S, Xiao J and Guo X (2005) Asymmetry analysis of cingulum based on scale-invariant parameterization by diffusion tensor imaging. *Hum Brain Mapp* 24(2):92–98.
- Good CD, Johnsrude I, Ashburner J, Henson RN, Friston KJ and Frackowiak RS (2001) Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *Neuroimage* 14(3):685–700.
- Gordh TE, Mair WGP and Sourander P (2007) Swedenborg, Linnaeus and brain research – and the roles of Gustaf Retzius and Alfred Stroh in the rediscovery of Swedenborg’s manuscripts. *Ups J Med Sci* 112(2):143–164.
- Gourine AV, Kasymov V, Marina N, Tang F, Figueiredo MF, Lane S, Teschemacher AG, Spyer KM, Deisseroth K and Kasparov S (2010) Astrocytes control breathing through pH-dependent release of ATP. *Science* 329(5991):571–575.
- Gross CG (1999) *Brain, Vision, Memory: Tales in the History of Neuroscience*. A Bradford Book. Cambridge: MIT Press, 273 pages.
- Gschmeissner S (2011) Nerve fibre, SEM. Science Photo Library, accessed on the 26th of August 2012.
URL: www.sciencephoto.com/media/307359/enlarge
- Guerrini A (2003) *Experimenting with Humans and Animals: From Galen to Animal Rights*. The Johns Hopkins University Press, 184 pages.
- Guevara P, Duclap D, Poupon C, Marrakchi-Kacem L, Fillard P, Le Bihan D, Leboyer M, Houenou J and Mangin JF (2012) Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *Neuroimage* 61(4):1083–1099.
- Guevara P, Poupon C, Rivière D, Cointepas Y, Descoteaux M, Thirion B and Mangin JF (2011) Robust clustering of massive tractography datasets. *Neuroimage* 54(3):1975–1993.
- Guha S, Rastogi R and Shim K (2001) Cure: An Efficient Clustering Algorithm for Large Databases. *Inf Syst* 26(1):35–58.
- Güllmar D, Langner J, Haueisen J and Reichenbach JR (2008) Combination of distance measures for optimal fiber clustering in diffusion tensor imaging. In: *Proc Intl Soc Mag Reson Med*, volume 16, Toronto, Canada.
- Güntürkün O (2005) How asymmetry in animals starts. *Eur Rev* 13(S2):105–118.
- Guo S, Kendrick KM, Zhang J, Broome M, Yu R, Liu Z and Feng J (2013) Brain-wide functional inter-hemispheric disconnection is a potential biomarker for schizophrenia and distinguishes it from depression. *Neuroimage Clin* 2:818–826.

- Haacke EM, Brown RW, Thompson MR and Venkatesan R (1999) *Magnetic Resonance Imaging Physical Principles and Sequence Design*. Wiley-Liss, 914 pages.
- Hansen P and Jaumard B (1997) Cluster analysis and mathematical programming. *Math Program* 79:191–215, 10.1007/BFo2614317.
- Heid O (2000) Eddy Current-Nulled Diffusion Weighting. In: *Proc Intl Soc Mag Reson Med*, 8, Denver, Colorado, USA.
- Heiervang E, Behrens TEJ, Mackay CE, Robson MD and Johansen-Berg H (2006) Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage* 33(3):867–877.
- Hellige J (1993) *Hemispheric Asymmetry: What's Right and What's Left*. Harvard University Press.
- Hendee WR and Ritenour ER (2002) *Medical Imaging Physics*. Wiley-Liss, 536 pages.
- Highley JR, Walker MA, Esiri MM, Crow TJ and Harrison PJ (2002) Asymmetry of the uncinate fasciculus: a post-mortem study of normal subjects and patients with schizophrenia. *Cereb Cortex* 12(11):1218–1224.
- Howes OD and Kapur S (2009) The dopamine hypothesis of schizophrenia: version III—the final common pathway. *Schizophr Bull* 35(3):549–562.
- Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, Calabresi PA, Pekar JJ, van Zijl PCM and Mori S (2008) Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage* 39(1):336–347.
- Huang H, Zhang J, Jiang H, Wakana S, Poetscher L, Miller MI, van Zijl PCM, Hillis AE, Wytik R and Mori S (2005) DTI tractography based parcellation of white matter: application to the mid-sagittal morphology of corpus callosum. *Neuroimage* 26(1):195–205.
- Hugdahl K (2005) Symmetry and asymmetry in the human brain. *Eur Rev* 13(S2):119–133.
- Hutter H (2008) What is *C. elegans*? Simon Fraser University, accessed on the 10th of September 2012.
URL: http://www.sfu.ca/biology/faculty/hutter/hutterlab/research/Ce_nervous_system.html
- Insel TR (2010) Rethinking schizophrenia. *Nature* 468(7321):187–193.
- Jarosz Q (2009) Structure of a Neuron. Wikipedia.
URL: http://en.wikipedia.org/wiki/File:Neuron_Hand-tuned.svg
- Javitt DC (2010) Glutamatergic theories of schizophrenia. *Isr J Psychiatry Relat Sci* 47(1):4–16.
- Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW and Smith SM (2012) FSL. *Neuroimage* 62(2):782–790.

Bibliography

- Jeong HK and Anderson AW (2008) Characterizing fiber directional uncertainty in diffusion tensor MRI. *Magn Reson Med* 60(6):1408–1421.
- Jékely G (2011) Origin and early evolution of neural circuits for the control of ciliary locomotion. *Proc Biol Sci* 278(1707):914–922.
- Jones DK, editor (2011) *Diffusion MRI: Theory, Methods and Applications*. Oxford Univ. Press, New York ; Oxford, 784 pages.
- Jones DK and Pierpaoli C (2005) Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. *Magn Reson Med* 53(5):1143–1149.
- Jones DK, Symms MR, Cercignani M and Howard RJ (2005) The effect of filter size on VBM analyses of DT-MRI data. *Neuroimage* 26(2):546–554.
- Kaiser M and Varier S (2011) Evolution and development of brain networks: from *Caenorhabditis elegans* to *Homo sapiens*. *Network* 22(1-4):143–147.
- Kanaan RA, Shergill SS, Barker GJ, Catani M, Ng VW, Howard R, McGuire PK and Jones DK (2006) Tract-specific anisotropy measurements in diffusion tensor imaging. *Psychiatry Res* 146(1):73–82.
- Kanaan RAA, Kim JS, Kaufmann WE, Pearlson GD, Barker GJ and McGuire PK (2005) Diffusion tensor imaging in schizophrenia. *Biol Psychiatry* 58(12):921–929.
- Karlsson B (2005) *Beyond the C++ Standard Library – An Introduction to Boost*. Addison-Wesley Longman, Amsterdam, 432 pages.
- Karypis G, Eui and News VK (1999) Chameleon: Hierarchical Clustering Using Dynamic Modeling. *Computer* 32(8):68–75.
- Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ and Parsey RV (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46(3):786–802.
- Koay CG, Nevo U, Chang LC, Pierpaoli C and Basser PJ (2008) The elliptical cone of uncertainty and its normalized measures in diffusion tensor imaging. *IEEE Trans Med Imaging* 27(6):834–846.
- Kovalev VA, Kruggel F and von Cramon DY (2003) Gender and age effects in structural brain asymmetry as measured by MRI texture analysis. *Neuroimage* 19(3):895–905.
- Kubicki M, McCarley R, Westin CF, Park HJ, Maier S, Kikinis R, Jolesz FA and Shenton ME (2007) A review of diffusion tensor imaging studies in schizophrenia. *J Psychiatr Res* 41(1-2):15–30.

- Lanyon LJ (2012) Diffusion Tensor Imaging: Structural Connectivity Insights, Limitations and Future Directions. In: *Neuroimaging – Methods*. InTech, 137–162.
- Lauterbur PC (1973) Image formation by induced local interactions. Examples employing nuclear magnetic resonance. *Nature* 242:190–191.
- Lawes INC, Barrick TR, Murugam V, Spierings N, Evans DR, Song M and Clark CA (2008) Atlas-based segmentation of white matter tracts of the human brain using diffusion tensor tractography and comparison with classical dissection. *Neuroimage* 39(1):62–79.
- Lawes INC and Clark CA (2011) Anatomical Validation of DTI and Tractography. In: *Diffusion MRI: Theory, Methods and Applications*. Oxford Univ. Press, New York ; Oxford, 439–447.
- Lazar M and Alexander AL (2005) Bootstrap white matter tractography (BOOT-TRAC). *Neuroimage* 24(2):524–532.
- Löbel U (2007) Diffusion Tensor Imaging: Darstellung der physiologischen Entwicklung des menschlichen Gehirns anhand des Diffusionskoeffizienten ADC und der Fraktionellen Anisotropie FA. Ph.D. thesis, Friedrich-Schiller-University.
- Le Bihan D and Breton E (1985) Imagerie de diffusion in vivo par résonance magnétique nucléaire. *C R Acad Sci Paris* 301:1109–1112.
- Le Bihan D, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N and Chabriat H (2001) Diffusion tensor imaging: Concepts and applications. *J Magn Reson Imaging* 13(4):534–546.
- Le Bihan D, Poupon C, Amadon A and Lethimonnier F (2006) Artifacts and pitfalls in diffusion MRI. *J Magn Reson Imaging* 24(3):478–488.
- Leemans A (2006) Modeling and Processing of Diffusion Tensor Magnetic Resonance Images for Improved Analysis of Brain Connectivity. Ph.D. thesis, Universiteit Antwerpen.
- Leergaard TB, White NS, de Crespigny A, Bolstad I, D’Arceuil H, Bjaalie JG and Dale AM (2010) Quantitative histological validation of diffusion MRI fiber orientation distributions in the rat brain. *PLoS ONE* 5(1):e8595.
- Levitt MH (2001) *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. Wiley-VCH, 710 pages.
- Li H, Xue Z, Guo L, Liu T, Hunter J and Wong STC (2010) A hybrid approach to automatic clustering of white matter fibers. *Neuroimage* 49(2):1249–1258.
- MacDonald AW and Schulz SC (2009) What we know: findings that every theory of schizophrenia should explain. *Schizophr Bull* 35(3):493–508.
- Maddah M (2008) Quantitative Analysis of Cerebral White Matter Anatomy from Diffusion MRI. Ph.D. thesis, Massachusetts Institute of Technology.

Bibliography

- Maddah M, Grimson WEL, Warfield SK and Wells WM (2008) A unified framework for clustering and quantitative analysis of white matter fiber tracts. *Med Image Anal* 12(2):191–202.
- Medkour T, Walden AT, Burgess AP and Strelets VB (2010) Brain connectivity in positive and negative syndrome schizophrenia. *Neuroscience* 169(4):1779–1788.
- Mentzel HJ, Ros C and Güllmar D (2011) [Diffusion magnetic resonance tomography of the brain in childhood]. *Monatsschr Kinderh* 159(4):325–332.
- Miller R (1996) *Axonal Conduction Time and Human Cerebral Laterality: A Psychological Theory*. CRC Press.
- Roberts B, Vilanova A and van Wijk JJ (2005) Evaluation of Fiber Clustering Methods for Diffusion Tensor Imaging. In: *VIS 05. IEEE*, IEEE Computer Society, 9.
- Mori S, Crain BJ, Chacko VP and van Zijl PC (1999) Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann Neurol* 45(2):265–269.
- Mori S and van Zijl PCM (2002) Fiber tracking: principles and strategies - a technical review. *NMR Biomed* 15(7-8):468–480.
- Moseley M (2002) Diffusion tensor imaging and aging - a review. *NMR Biomed* 15(7-8):553–560.
- Nicholson C (2001) Diffusion and related transport mechanisms in brain tissue. *Rep Prog Phys* 64(7):815.
- Nucifora PGP, Verma R, Lee SK and Melhem ER (2007) Diffusion-tensor MR imaging and tractography: exploring brain microstructure and connectivity. *Radiology* 245(2):367–384.
- NVIDIA (2010) *NVIDIA CUDA C Programming Guide*. Version 3.1.1.
- Object Management Group (2011) OMG Unified Modeling Language (OMG UML), Infrastructure, V2.4.1. Technical report.
- O'Donnell LJ, Kubicki M, Shenton ME, Dreusicke MH, Grimson WEL and Westin CF (2006) A method for clustering white matter fiber tracts. *AJNR Am J Neuroradiol* 27(5):1032–1036.
- O'Donnell LJ and Westin CF (2007) Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans Med Imaging* 26(11):1562–1575.
- O'Donnell LJ, Westin CF and Golby AJ (2007) Tract-based morphometry. *Med Image Comput Comput Assist Interv* 10(Pt 2):161–168.
- O'Donnell LJ, Westin CF and Golby AJ (2009) Tract-based morphometry for white matter group analysis. *Neuroimage* 45(3):832–844.

- Owen MJ, O'Donovan MC, Thapar A and Craddock N (2011) Neurodevelopmental hypothesis of schizophrenia. *Br J Psychiatry* 198(3):173–175.
- Owens JD, Houston M, Luebke D, Green S, Stone JE and Phillips JC (2008) GPU Computing. *Proc IEEE* 96(5):879–899.
- Ozturk A, Sasson AD, Farrell JAD, Landman BA, da Motta ACBS, Aralasmak A and Yousem DM (2008) Regional differences in diffusion tensor imaging measurements: assessment of intrarater and interrater variability. *AJNR Am J Neuroradiol* 29(6):1124–1127.
- Pandit L (2009) Differential diagnosis of white matter diseases in the tropics: An overview. *Ann Indian Acad Neurol* 12(1):12–21.
- Park HJ, Westin CF, Kubicki M, Maier SE, Niznikiewicz M, Baer A, Frumin M, Kikinis R, Jolesz FA, McCarley RW and Shenton ME (2004) White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study. *Neuroimage* 23(1):213–223.
- Parker GJM (2011) Probabilistic Fiber Tracking. In: *Diffusion MRI: Theory, Methods and Applications*. Oxford Univ. Press, New York ; Oxford, 395–408.
- Parker GJM and Alexander DC (2005) Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Philos Trans R Soc Lond B Biol Sci* 360(1457):893–902.
- Partadiredja G, Miller R and Oorschot DE (2003) The number, size, and type of axons in rat subcortical white matter on left and right sides: a stereological, ultrastructural study. *J Neurocytol* 32(9):1165–1179.
- Perrin J (1909) Mouvement brownien et réalité moléculaire. *Ann Ch Ph* 18:5–104.
- Picchioni MM and Murray RM (2007) Schizophrenia. *BMJ* 335(7610):91–95.
- Pierpaoli C and Basser PJ (1996) Toward a quantitative assessment of diffusion anisotropy. *Magn Reson Med* 36(6):893–906.
- Price WS (1997) Pulsed-field Gradient Nuclear Magnetic Resonance As a Tool for Studying Translational Diffusion, Part 1: Basic Theory. *Magn Reson Educ J* 9(5):299–336.
- Purcell EM, Torrey HC and Pound RV (1946) Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys Rev* 69(1-2):37–38.
- Ramón y Cajal S (1913) *Estudios sobre la degeneración y regeneración del sistema nervioso*.
- Ratcliff WC, Denison RF, Borrello M and Travisano M (2012) Experimental evolution of multicellularity. *Proc Natl Acad Sci USA* 109(5):1595–1600.

Bibliography

- Ros C (2007) Rekonstruktion von MR Magnituden- und Phasenbildern aus Multikanal-spulen unter Berücksichtigung der Spulensensitivitäten. Master's thesis, Friedrich-Schiller-University.
- Rucklidge W (1996) Efficient visual recognition using the Hausdorff distance. In: *Lecture Notes in Computer Science*, volume 1173, Springer Heidelberg.
- Sarnat HB and Netsky MG (2002) When does a ganglion become a brain? Evolutionary origin of the central nervous system. *Semin Pediatr Neurol* 9(4):240–253.
- Schweser F, Deistung A, Lehr BW and Reichenbach JR (2011) Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: an approach to in vivo brain iron metabolism? *Neuroimage* 54(4):2789–2807.
- Schweser F, Sommer K, Deistung A and Reichenbach JR (2012) Quantitative susceptibility mapping for investigating subtle susceptibility variations in the human brain. *Neuroimage* 62(3):2083–2100.
- Sharan R and Shamir R (2000) CLICK: A Clustering Algorithm with Applications to Gene Expression Analysis. In: *Proc Int Conf Intell Syst Mol Biol*, 8, AAAI Press, 307–316.
- Shenton ME, Dickey CC, Frumin M and McCarley RW (2001) A review of MRI findings in schizophrenia. *Schizophr Res* 49(1-2):1–52.
- Shimony JS, Snyder AZ, Lori N and Contum TE (2002) Automated Fuzzy Clustering of Neuronal Pathways in Diffusion Tensor Tracking. In: *Proc Intl Soc Mag Reson Med*, 10, Honolulu, Hawaii, USA.
- Skopal T and Bustos B (2011) On nonmetric similarity search problems in complex domains. *ACM Comput Surv* 43(4):34:1–34:50.
- Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM and Behrens TEJ (2006) Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31(4):1487–1505.
- Sperry RW (1961) Cerebral Organization and Behavior: The split brain behaves in many respects like two separate brains, providing new research possibilities. *Science* 133(3466):1749–1757.
- Stadlbauer A, Nimsky C, Buslei R, Salomonowitz E, Hammen T, Buchfelder M, Moser E, Ernst-Stecken A and Ganslandt O (2007) Diffusion tensor imaging and optimized fiber tracking in glioma patients: Histopathologic evaluation of tumor-invaded white matter structures. *Neuroimage* 34(3):949–956.
- Stejskal EO and Tanner JE (1965) Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *J Chem Phys* 42(1):288–292.

- Stephan KE, Friston KJ and Frith CD (2009) Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophr Bull* 35(3):509–527.
- Syková E and Nicholson C (2008) Diffusion in brain extracellular space. *Physiol Rev* 88(4):1277–1340.
- Takahashi M, Hackney DB, Zhang G, Wehrli SL, Wright AC, O'Brien WT, Uematsu H, Wehrli FW and Selzer ME (2002) Magnetic resonance microimaging of intraaxonal water diffusion in live excised lamprey spinal cord. *Proc Natl Acad Sci USA* 99(25):16192–16196.
- Takao H, Hayashi N and Ohtomo K (2011) White matter asymmetry in healthy individuals: a diffusion tensor imaging study using tract-based spatial statistics. *Neuroscience* 193:291–299.
- Tamraz JC and Comair YG (2005) *Atlas of Regional Anatomy of the Brain Using MRI: With Functional Correlations*. Springer Berlin, 2nd edition, 345 pages.
- Thiebaut de Schotten M, Ffytche DH, Bizzi A, Dell'Acqua F, Allin M, Walshe M, Murray R, Williams SC, Murphy DGM and Catani M (2011) Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *Neuroimage* 54(1):49–59.
- Todd K, Serrano A, Lacaille J and Robitaille R (2006) Glial cells in synaptic plasticity. *J Physiol Paris* 99(2-3):75–83.
- Toga AW and Thompson PM (2003) Mapping brain asymmetry. *Nat Rev Neurosci* 4(1):37–48.
- Tuch DS (2004) Q-ball imaging. *Magn Reson Med* 52(6):1358–1372.
- Twieg DB (1983) The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Med Phys* 10(5):610–621.
- U S National Library of Medicine (2012) The Edwin Smith Surgical Papyrus. Internet. URL: archive.nlm.nih.gov/proj/ttp/books.htm
- Ursell TS (2011) The Diffusion Equation: A Multi-dimensional Tutorial. Technical report, Department of Applied Physics; California Institute of Technology.
- van Os J and Kapur S (2009) Schizophrenia. *Lancet* 374(9690):635–645.
- Van Wagenen WP and Herren RY (1940) Surgical division of commissural pathways in the corpus callosum: relation to spread of an epileptic attack. *Archives of Neurology and Psychiatry* 44(4):740.
- Verhoeven JS, Sage CA, Leemans A, Hecke WV, Callaert D, Peeters R, Cock PD, Lagae L and Sunaert S (2010) Construction of a stereotaxic DTI atlas with full diffusion tensor information for studying white matter maturation from childhood to adolescence using tractography-based segmentations. *Hum Brain Mapp* 31(3):470–486.

Bibliography

- Visser E, Nijhuis EHJ, Buitelaar JK and Zwiers MP (2011) Partition-based mass clustering of tractography streamlines. *Neuroimage* 54(1):303–312.
- von Neumann J (1945) First Draft of a Report on the EDVAC. Technical report, University of Pennsylvania.
- Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, Hua K, Zhang J, Jiang H, Dubey P, Blitz A, van Zijl P and Mori S (2007) Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 36(3):630–644.
- Wakana S, Jiang H, Nagae-Poetscher LM, van Zijl PCM and Mori S (2004) Fiber tract-based atlas of human white matter anatomy. *Radiology* 230(1):77–87.
- Wang Q, Yap PT, Wu G and Shen D (2013) Application of neuroanatomical features to tractography clustering. *Hum Brain Mapp* 34(9):2089–2102.
- Wang R, Benner T, Sorensen AG and Wedeen VJ (2007) Diffusion Toolkit: A Software Package for Diffusion Imaging Data Processing and Tractography. In: *Proc Intl Soc Mag Reson Med*, 15, Berlin, Germany.
- Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* 58(301):236–244.
- Wassermann D, Bloy L, Kanterakis E, Verma R and Deriche R (2010) Unsupervised white matter fiber clustering and tract probability map generation: applications of a Gaussian process framework for white matter fibers. *Neuroimage* 51(1):228–241.
- Weinberger D (1986) The pathogenesis of schizophrenia: a neurodevelopmental theory. *Handbook of schizophrenia* 1:397–406.
- Wernicke C (1874) *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Cohn.
- Wheeler-Kingshott CAM and Cercignani M (2009) About "axial" and "radial" diffusivities. *Magn Reson Med* 61(5):1255–1260.
- White JG, Southgate E, Thomson JN and Brenner S (1986) The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci* 314(1165):1–340.
- Whitford TJ, Kubicki M and Shenton ME (2011) Diffusion tensor imaging, structural connectivity, and schizophrenia. *Schizophr Res Treatment* 2011:709523.
- Williams TH, Gluhbegovic N and Jew JY (1999) The Human Brain: Dissections of the Real Brain. Internet.
URL: www.scribd.com/doc/28155705/The-Human-Brain-Dissections-of-the-Real-Brain
- Wu X, Xie M, Zhou J, Anderson AW, Gore JC and Ding Z (2012) Globally optimized fiber tracking and hierarchical clustering – a unified framework. *Magn Reson Imaging* 30(4):485–495.

- Xu R and Wunsch DC (2008) *Clustering*. Wiley-IEEE Press.
- Yeatman JD, Dougherty RF, Myall NJ, Wandell BA and Feldman HM (2012) Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS ONE* 7(11):e49790.
- Yendiki A, Panneck P, Srinivasan P, Stevens A, Zöllei L, Augustinack J, Wang R, Salat D, Ehrlich S, Behrens T, Jbabdi S, Gollub R and Fischl B (2011) Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Front Neuroinform* 5:23.
- Zhang S, Correia S and Laidlaw DH (2008) Identifying white-matter fiber bundles in DTI data using an automated proximity-based fiber-clustering method. *IEEE T Vis Comput Gr* 14(5):1044–1053.
- Zhang Y, Zhang J, Oishi K, Faria AV, Jiang H, Li X, Akhter K, Rosa-Neto P, Pike GB, Evans A, Toga AW, Woods R, Mazziotta JC, Miller MI, van Zijl PCM and Mori S (2010) Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *Neuroimage* 52(4):1289–1301.
- Zoli M, Torri C, Ferrari R, Jansson A, Zini I, Fuxe K and Agnati L (1998) The emergence of the volume transmission concept. *Brain Res Rev* 26(2):136–147.

List of Figures

1.1	Visualization of the nervous system of <i>Caenorhabditis elegans</i>	4
1.2	First occurrence of the word brain	5
1.3	Drawing of the brain and the ventricular system	6
1.4	Drawing that depicts the white and grey matter	6
1.5	Drawing that illustrates regeneration of a nerve fiber after hemisection . . .	7
1.6	Depiction of <i>post mortem</i> tracts and <i>in vivo</i> fiber tracts	8
2.1	Ellipsoidal representation of diffusion tensors	17
2.2	Simplified schematic of a neuron	18
2.3	Images of pyramidal neurons and a disconnected bundle of axons	20
2.4	Bohrs model of an atomic nucleus	21
2.5	Atomic nucleus with a nuclear spin in a static magnetic field B_0	21
2.6	A HF pulse tips the magnetization into the transverse plane	21
2.7	Example of a 2D k-space and the corresponding image of the spin density . .	22
2.8	Schematic illustration of the diffusion weighted PGSE experiment	24
2.9	Transverse MRI images with and without diffusion weighting	26
2.10	Diffusion tensors and various DTI contrasts	28
2.11	Illustration of tensor based fiber tractography for one tract	28
2.12	Examples of fiber tracking	29
2.13	Formal definition of a fiber tract	29
2.14	Illustration of the cluster analysis concept	33
2.15	Relationship between features and similarity measures	34
2.16	CUDA-enabled GPU architecture	38
3.1	Whole brain fiber tractography and fiber tract clustering	43
3.2	Overview of CATSER	46
3.3	Color-coded visualization of LOFs	49
3.4	Influence of LOFs on intra-cluster distances	50
3.5	Influence of LOFs on inter-cluster distances	51
3.6	Influence of the number of representatives on the clustering process	52
3.7	Determination of the number of representatives	53
3.8	Assignment and reassignment of tracts in CATSER	54
3.9	Effects of different weighting factors	58
3.10	A single fiber tract that traverses through two classes of an atlas	59

List of Figures

4.1	Illustration of a fiber tract and its segment centroids	64
4.2	Illustration of the MPD	67
4.3	Illustration of sub-stages for the computation of the CD	69
5.1	The stages of the clustering and their sequential and parallel sub-parts	74
5.2	Concurrent computation of similarity measures between two sets of tracts .	77
5.3	Concurrent computation of similarity measures for one set of tracts	77
5.4	Extended relationship between features and similarity measures	78
5.5	Example for the implementation of similarity measures with and without result reuse	79
5.6	Flowchart for determining the processing order of features and similarity measures	80
5.7	Operator view and internal view for two regular similarity measures	81
5.8	Allocated memory used during the computation of features and similarities for two sets of tracts	82
5.9	Allocated memory used during the computation of features and similarity matrices for one set of tracts	83
6.1	Example for an atlas class in the probabilistic white matter atlas	90
6.2	Volume renderings for a selection of fiber bundles of the white matter atlas .	91
6.3	Atlas-guided clustering results for one dataset, clustered with CD	95
6.4	Spatial agreement of clustered fiber bundles and atlas classes for CD	96
6.5	Spatial agreement of clustered fiber bundles and atlas classes for HD	97
6.6	Spatial agreement of clustered fiber bundles and atlas classes for MPD	98
6.7	IFO of one dataset clustered with different methods	100
6.8	Left UNC of two datasets clustered with atlas-guided CATSER	101
6.9	Left SLFt of 15 datasets	102
6.10	Results to demonstrate the effects of outlier elimination	103
6.11	Performance analysis of the clustering with CD, HD and MPD	104
6.12	Computation times and speedup for HD with a CPU system	105
6.13	Computation times for HD with CPU and GPU systems	105
7.1	Illustration of normalization-induced interpolation	111
7.2	Illustration of fiber bundle-based processing that avoids normalization- induced interpolation	112
7.3	Workflow of the quantitative fiber bundle-driven analysis	113
7.4	Projection of quantitative values onto tract points	114
7.5	Gridding of quantitative values	116
7.6	Number of datasets that contribute to the statistical analysis of the SLFt . . .	117
7.7	Overlapping regions of the IFO and the UNC bundle	120
7.8	Average difference per voxel in dependence of resolution and number of tracts	122
7.9	Average difference per voxel in dependence of resolution and tract point distance	123
7.10	Distribution of values after gridding of fiber bundles with different methods	124

7.11	Distributions of voxel value differences between conventional and fiber bundle-based gridding	126
8.1	Hemispheric differences in the CGC of healthy volunteers	132
8.2	Hemispheric differences in the UNC of healthy volunteers	133
8.3	Hemispheric differences in the CST of healthy volunteers	134
8.4	Hemispheric differences in the IFO of healthy volunteers	134
8.5	Hemispheric differences in the SLFt of healthy volunteers	135
8.6	The segments of the arcuate fasciculus	135
8.7	Hemispheric differences in the CGC of schizophrenic patients	139
8.8	Hemispheric differences in the UNC of schizophrenic patients	139
8.9	Hemispheric differences in the CST of schizophrenic patients	140
8.10	Hemispheric differences in the IFO of schizophrenic patients	140
8.11	Hemispheric differences in the SLFt of schizophrenic patients	141

List of Tables

2.1	$N \times N$ similarity matrix \mathcal{S}	35
2.2	Flynn's taxonomy	36
3.1	Overview of clustering parameters	56
3.2	The four cases that determine the weighting factor for the atlas guidance . .	61
6.1	Parameters for the different outlier elimination strategies	93
6.2	Technical details for the employed CPU and GPU systems	94
6.3	Average spatial agreement and standard deviation for all clustering techniques and similarity measures	99
7.1	Parameters for the investigation of the dependency between resolution, num- ber of tracts and distances between tract points	118
8.1	Summary of differences in diffusion parameters for statistically significant regions of bundles from the left and right hemisphere in healthy volunteers .	132
8.2	Summary of differences in diffusion parameters for statistically significant regions of bundles from the left and right hemisphere in schizophrenic patients	137

Part V

Appendix

Publications

Journal papers

- Ros C**, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2014) A new fiber bundle-driven method for the quantitative analysis of diffusion-weighted MRI data and its application to assess hemispheric differences in fiber bundles. *Manuscript in preparation*.
- Ros C**, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2013) Atlas-guided cluster analysis of large tractography datasets. *PLoS ONE* 8(12):e83847.
- Hiepe P, Herrmann KH, Güllmar D, **Ros C**, Siebert T, Blickhan R, Hahn K and Reichenbach JR (2013) Fast low-angle shot diffusion tensor imaging with stimulated echo encoding in the muscle of rabbit shank. *NMR Biomed*. Epub ahead of print.
- Ros C**, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2012) Quantitative fiber bundle-based analysis of diffusion-weighted MRI data. *Biomed Tech (Berl)* 57:530–533.
- Mentzel HJ, **Ros C** and Güllmar D (2011) [Diffusion magnetic resonance tomography of the brain in childhood]. *Monatsschr Kinderheilkd* 159:325–332, 10.1007/s00112-010-2319-9.
- Hiepe P, Herrmann KH, **Ros C** and Reichenbach JR (2011) Diffusion weighted inner volume imaging of lumbar disks based on turbo-STEAM acquisition. *Z Med Phys* 21(3):216–227.
- Roßbach M, Jochimsen TH, Herrmann KH, **Ros C**, Güllmar D and Reichenbach JR (2011) [Comparison of long-axis and short-axis PROPELLER-EPI for diffusion-weighted Magnetic Resonance Imaging]. *Z Med Phys* 21(1):33–41.
- Ros C**, Güllmar D, Klatt M and Reichenbach JR (2009) Visualization of Orientation Distribution Functions (ODFs) with MATLAB. In: Dössel O and Schlegel WC, editors, *World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings*, volume 25, Springer-Verlag GmbH, ISBN: 978-3-642-03903-4, 248–251.
- Ros C**, Witoszynskyj S, Herrmann KH and Reichenbach JR (2009) Reconstruction of phase images for GRAPPA accelerated Magnetic Resonance Imaging. In: Sloten J, Verdonck P, Nyssen M and Haueisen J, editors, *4th European Conference of the International Federation for Medical and Biological Engineering ECIFMBE 2008, IFMBE Proceedings*, volume 22, Springer-Verlag GmbH, ISBN: 978-3-540-89207-6, 803–806.
- Hiepe P, **Ros C**, Reichenbach JR and Herrmann KH (2009) Diffusion Weighted ZOOM Imaging in the Lumbar Spine Based on Single-Shot STEAM. In: Dössel O and Schlegel WC,

editors, *World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings*, volume 25, Springer-Verlag GmbH, ISBN: 978-3-642-03878-5, 670–672.

Conference papers

Ros C, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2013) Quantitative fiber bundle-driven analysis of diffusion MRI data. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 21.

Ros C, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2013) Atlas-guided cluster analysis of fiber tracts. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 21.

Al-Shuhari F, Güllmar D, Löbel U, **Ros C**, Husain R, Reichenbach JR, Stenzel M and Mentzel HJ (2013) Diffusion tensor imaging (DTI) bei Kindern und Jugendlichen mit Neurofibromatose Typ I. In: *Jahrestagung der Gesellschaft für Pädiatrische Radiologie (GPR)*, volume 50.

Ros C, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2012) A semi-automatic approach for the extraction of white matter fiber bundles across subjects. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 20.

Ros C, Muller-Moran HR, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2012) Tractography-based method for the estimation of the co-registration error for white matter structures. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 20.

Ros C, Güllmar D, Stenzel M, Reichenbach JR and Mentzel HJ (2012) COMET (Cluster analysis Of Major Equivalent Tracts) - A framework for the automatic bundling of fiber tracts. In: *European Congress of Radiology*, Vienna, Austria.

Ros C, Güllmar D, Stenzel M, Reichenbach JR and Mentzel HJ (2012) Semi-automatische Extraktion von Faserbündeln aus multiplen MR-DTI Datensätzen. In: *Röfo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 93.

Ros C, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2012) Quantitative fibre bundle-based analysis of diffusion-weighted MRI data. In: *Jahrestagung der Deutschen Gesellschaft für Medizinische Physik*, volume 43, Jena, Germany.

Ros C, Muller-Moran HR, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2012) Tractography-based method for the estimation of the co-registration error for white matter structures. In: *Jahrestagung der Deutschen Gesellschaft für Medizinische Physik*, volume 43, Jena, Germany.

Ros C, Güllmar D, Stenzel M, Mentzel HJ and Reichenbach JR (2012) A semi-automatic approach for the extraction of white matter fiber bundles across subjects. In: *Jahrestagung der Deutschen Gesellschaft für Medizinische Physik*, volume 43, Jena, Germany.

- Hiepe P, Güllmar D, **Ros C**, Siebert T, Gussew A, Rzanny R and Reichenbach JR (2012) Functional DTI in Voluntary Contracted Human Calf Muscles Using an MR Compatible Ergometer. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 20.
- Güllmar D, Lange C, **Ros C**, Deistung D and Reichenbach JR (2012) Using wild bootstrap to evaluate the effect of spatial resolution on MR diffusion parameters. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 20.
- Ros C**, Tandetzky R, Güllmar D and Reichenbach JR (2011) GPGPU-Computing for the cluster analysis of fiber tracts: Replacing a \$15000 high end PC with a \$500 graphics card. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 19.
- Ros C**, Güllmar D and Reichenbach JR (2011) COMET - A framework for the large scale Cluster analysis Of Major Equivalent Tracts. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 19.
- Ros C**, Güllmar D, Stenzel M, Reichenbach JR and Mentzel HJ (2011) Vollautomatische Extraktion von Faserbündeln aus DTI-Fibertracking-Datensätzen durch Anwendung von Clusteranalyseverfahren. In: *Jahrestagung der Deutschen Gesellschaft für Neuroradiologie*, volume 46.
- Ros C**, Güllmar D and Reichenbach JR (2011) COMET (Cluster analysis Of Major Equivalent Tracts) - Ein Framework für die vollautomatische Clusteranalyse von Fiber Tracts. In: *Röfo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 92.
- Güllmar D, **Ros C** and Reichenbach JR (2011) Effect of using super-resolution technique in slice direction on DTI fiber tractography. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 19.
- Hiepe P, Herrmann KH, **Ros C** and Reichenbach JR (2011) Diffusionsgewichtete MR-Bildgebung lumbaler Bandscheiben mittels Inner-Volume-STEAM-Technik. In: *Röfo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 92.
- Ros C**, Güllmar D and Reichenbach JR (2010) A new combined distance measure for the clustering of fiber tracts in DTI. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 18.
- Hiepe P, Herrmann KH, **Ros C** and Reichenbach JR (2010) Diffusion Weighted Turbo-STEAM ZOOM Imaging of the Lumbar Spine. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 18.
- Ros C**, Sedlacik J, Daw NC, Loeffler R, Reichenbach JR and Hillenbrand CM (2009) Boosting image quality by automatic low SNR channel detection and its removal in k-space. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 17.

- Henf AM, Herrmann KH, **Ros C** and Reichenbach JR (2009) Optimierte komplexwertige Keyhole-Rekonstruktion von MR-Daten. In: *Proceedings of the German Chapter of the International Society for Magnetic Resonance in Medicine*, volume 12.
- Ros C**, Witoszynskyj S, Herrmann KH and Reichenbach JR (2008) Reconstruction of Phase Images for GRAPPA based Susceptibility Weighted Imaging (SWI). In: *Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 16, Toronto, Canada.
- Ros C**, Witoszynskyj S, Herrmann KH and Reichenbach JR (2008) Rekonstruktion von Phasenbildern für die GRAPPA-basierte suszeptibilitätsgewichtete Bildgebung (SWI). In: *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 89, Berlin, Germany.
- Hiepe P, Herrmann KH, **Ros C**, Bellemann ME and Reichenbach JR (2008) Single-shot STEAM DWI: Erste in-vitro und in-vivo Ergebnisse. In: *Proceedings of the German Chapter of the International Society for Magnetic Resonance in Medicine*, volume 11, Frankfurt, Germany.
- Ros C**, Witoszynskyj S, Herrmann KH and Reichenbach JR (2007) Phasenrichtige Rekonstruktion von Multikanal MR-Bildern mit dem SENSE-Algorithmus für die suszeptibilitätsgewichtete Bildgebung (SWI). In: *Proceedings of the German Chapter of the International Society for Magnetic Resonance in Medicine*, volume 10, Würzburg, Germany.

Monographs and book contributions

- Ros C** (2007) Rekonstruktion von MR Magnituden- und Phasenbildern aus Multikanal-spulen unter Berücksichtigung der Spulensensitivitäten. Master's thesis, Friedrich-Schiller-University Jena.
- Ros C** (2006) File Transfer Protocol (FTP). In: Kauhaus C, editor, Prof. Dr.-Ing. W. Erhard, publisher, *Berichte zur Rechnerarchitektur* 12(11):99–110.

Invited talks

- Ros C** (2013) Clusteranalyse von Fibertracts und quantitative Faserbündel-gestützte Analysen der weißen Substanz. In: *Forschungsseminar am Institut für Biomedizinische Technik und Informatik der Technischen Universität Ilmenau*, Ilmenau, Germany.
- Ros C** (2010) Techniken zur Beschleunigung der k-Raum Akquise in der MRT. In: *Weiterbildungsveranstaltung am Universitätsklinikum Jena*, Jena, Germany.
- Ros C** (2009) Clustering von Fibertracts. In: *Bildgebende Verfahren in den Neurowissenschaften: Grundlagen und aktuelle Ergebnisse – Forschungsseminar am Universitätsklinikum Jena*, Jena, Germany.

Ros C (2008) Image reconstruction and processing of multi channel data in Magnetic Resonance Imaging. In: *North American IDEA Users Group Meeting*, Memphis, USA.

Ehrenwörtliche Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

1. *Unterstützung beim Einzeichnen von "regions of interest" (ROIs) und bei der Extraktion der dazugehörigen Fibertracts für die Erstellung des Atlases der weißen Substanz (vgl. Abschnitt 6.2).*

Daniel Güllmar, Dr.-Ing.

Medizinische Physik

Institut für Diagnostische und Interventionelle Radiologie I

Zentrum für Radiologie

Universitätsklinikum Jena, Jena, Deutschland

2. *Unterstützung bei der Implementierung der Hausdorff Distanz für die CUDA Plattform und Optimierung des bestehenden Programmcodes für die Berechnung der Hausdorff Distanz auf CPU Systemen (vgl. Abschnitt 5.3).*

Ralph Tandetzky, Dipl.-Math.

ehemaliger Mitarbeiter der Medizinischen Physik

Institut für Diagnostische und Interventionelle Radiologie I

Zentrum für Radiologie

Universitätsklinikum Jena, Jena, Deutschland

3. *Akquise von Probanden und Patienten sowie Organisation und Durchführung der MR-Messungen (vgl. Abschnitte 6.1 und 8.2.2).*

Mitarbeiter der Klinik für Psychiatrie und Psychotherapie

Universitätsklinikum Jena, Jena, Deutschland

sowie

Mitarbeiter des Instituts für Psychologie

Friedrich-Schiller-Universität Jena, Jena, Deutschland

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalte der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch bewertet wird und gemäß §7 Abs. 8 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat.

Jena, den 13. Januar 2014

(Ort, Datum)

.....

(Unterschrift)