TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Informatik
Computer Aided Medical Procedures & Augmented Reality / I16

Image Analysis Methods in Multi-View Transcranial 3D Ultrasound for Diagnosis of Neurological Movement-Disorders

Seyed-Ahmadi Ahmadi

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Hans-Joachim Bungartz
Prüfer der Dissertation:
1. Univ.-Prof. Dr. Nassir Navab
2. Prof. Benoit M. Dawant, Ph.D., Vanderbilt University, USA

Die Dissertation wurde am 26.09.2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 04.04.2013 angenommen.

Abstract

In the past two decades, transcranial ultrasound (TCUS), i.e. non-invasive B-mode imaging through the temporal bone window of the skull, has been established as a technique for diagnosis of neurological movement disorders such as Parkinson's Disease (PD). Although the TCUS method bears high potential, especially for early diagnosis and screening of PD, it is also critized since it requires a lot of experience from the sonographer, e.g. for finding a suitable cut-plane through the midbrain.

In this thesis, we pursue a paradigm shift of the TCUS method from regular 2D ultrasound to 3D-TCUS. We argue that the extension to 3D makes the technique more objective and facilitates acquisition for less experienced TCUS sonographers. It also allows for volumetric analysis of diagnostically relevant substantia nigra hyper-echogenicities (SNE) for the first time in this field.

As a first contribution, we propose to acquire 3DUS bi-laterally through both temporal bone windows, in order to maximize the available information on the target region. This acquisition is affected by linear and non-linear distortions, e.g. due to mis-calibration and speed-of-sound deviations in bone. We therefore formulate this scenario as a general multi-view 3DUS reconstruction problem, where each view is affected by diffeomorphic, non-linear distortions unknown during acquisition. For simultaneous recovery and compensation in an arbitrary number of 3D views, we introduce a novel compounding scheme, which performs several inter-dependent segmentation, registration and reconstruction steps to reach an optimal joint compounding. A statistical shape model (SSM) aids as a geometric prior and regularizes the reconstruction by constraining deformation fields within the manifold of legal anatomic shapes.

For segmentation of anatomic target regions, we propose two novel approaches for general detection of surfaces and lesions in 3DUS data, which are adapted to the local echo responses caused by US physics. The methods include 1) a (semi-)automatic segmentation in 3D-TCUS using an explicit, SSM-regularized active surface method with a localized region-based cost function and 2) an automatic detection of hyper-echogenic lesions using a probabilistic Random Forest classification, which generates fuzzy posteriors for 3DUS voxels based on descriptors of local intensity and contrast. We furthermore introduce a novel approach for the formulation of a Random Forest spatial prior which can be jointly learned during the Random Forest training stage.

We demonstrate the translational relevance of our approaches for early diagnosis of PD using an unprecendented dataset of 22 subjects, with 11 diagnosed PD patients and 11 healthy controls. Based on feature sets extracted from volumetric segmentations, we report on first, promising results of automatically distinguishing healthy controls from PD patients using Support Vector Machine classification, with sensitivities and specificities up to 90.9% and 72.3%, respectively. We believe that our findings mark the first steps towards an objective, robust and routinely applicable computer-aided system for early diagnosis of PD in future.

Zusammenfassung

In den letzten zwei Jahrzehnten hat sich transkranieller Ultraschall (TKUS), d.h. Ultraschall durch das temporale Knochenfenster des Schädels, als Methode zur Diagnostik neurologischer Bewegungsstörungen wie dem Parkinson Syndrome (PS) etabliert. Trotz vielversprechender Ergebnisse, insbesondere für Früh-Diagnostik und zur Vorsorge-Untersuchung, wird die TKUS-Methode kritisiert, da die Untersuchung viel Erfahrung benötigt, z.B. zur Ermittlung der optimalen Schnittebene durch das Mittelhirn.

In dieser Doktorarbeit verfolgen wir einen Paradigmenwechsel der TKUS Methode von regulärem 2D Ultraschall zu einem 3D Ansatz, der die Methode objektiver werden lässt und die Akquise auch für weniger erfahrene TKUS Untersucher erleichtert. Darüberhinaus erlaubt eine 3D Akquise zum ersten Mal eine volumetrische Analyse diagnostisch relevanter Hyper-Echogenitäten der Substantia Nigra (SNE).

Um die Bild-Information in der Zielregion zu maximieren, führen wir zunächst eine bi-laterale 3D Daten-Akquise durch beide temporalen Knochenfenster ein. Derlei Aufnahmen sind mit linearen und nicht-linearen Verzerrungen behaftet, z.B. aufgrund von Fehl-Kalibrierungen oder Abweichungen der Schallgeschwindigkeit in Knochengewebe. Wir formulieren diese Aufnahmetechnik daher allgemein als ein multi-laterales 3DUS Rekonstruktions-Problem, in dem jede Ansicht durch ein diffeomorphes, nicht-lineares Feld verzerrt ist, welches zum Zeitpunkt der Aufnahme unbekannt ist. Für eine neuartige, gleichzeitige Kompensation in einer beliebigen Anzahl an 3D Ansichten, führen wir mehrere voneinander abhängige Segmentierungs-, Registrierungs- und Rekonstruktionsschritte durch, um eine optimale Konsensus-Rekonstruktion zu erzielen. Ein statistisches Form-Modell (SFM) dient als geometrische Regularisierung, die die Deformationsfelder während der Rekonstruktion auf anatomisch sinnvolle Formen beschränkt.

Für die Segmentierung der anatomischen Zielregionen entwickeln wir zwei neuartige Ansätze zur generellen Detektion von Oberflächen und Läsionen in lokal-echogenen 3DUS Daten. Diese Methoden beinhalten 1) eine semi-automatische Segmentierung mittels einer expliziten, SFM-regularisierten Aktive-Oberflächen-Methode mit einer lokalisierten Kostenfunktion und 2) eine automatische Detektion von hyper-echogenen Läsionen mittels einer probabilistischen Random-Forest Klassifizierung, die auf lokalen Intensitäts- und Kontrast-Deskriptoren basiert. Wir präsentieren zudem einen neuartigen Ansatz, um Vorwissen über räumliche Verteilung von SNEs in der Lernphase des Random-Forest Klassifizierers mit einzubeziehen.

Wir demonstrieren die Relevanz unserer Methoden für die Parkinson-Frühdiagnostik erstmalig anhand eines klinischen Datensatzes mit 11 diagnostizierten Parkinson Patienten und 11 gesunden Kontrollprobanden. Mittels volumetrischer Merkmale aus den 3D Segmentierungen und Support-Vector-Machines präsentieren wir vielversprechende Ergebnisse zur automatischen Klassifizierung von PS Patienten und gesunden Kontrollprobanden. Sensitivität und Spezifitäten erreichen jeweils bis zu 90.9% und 72.3%. Wir sehen dies als einen ersten Schritt zu einem objektiveren, robusten und routinemäßig anwendbaren, computer-assistierten Diagnosesystem für PS in Zukunft.

Acknowledgments

After four and a half years of research for this thesis, and after close to eight years of active collaboration at the CAMP chair...

...my first and foremost thanks go to Prof. Nassir Navab, who has very early on recognized and nourished my fascination with medical technologies and allowed me to slowly but gradually become a defector from the Electrical Engineering to the Computer Science faculty. I cannot find enough words of appraisal for his kind guidance, his enthusiastic support, innovativeness and patience throughout the past years. Thanks for eight years of fun!

My next thanks go to a set of wonderful colleagues and contributors to this thesis, most of all Tassilo Klein, with whom I have had the joy and honor of sharing a quasi-Siamese research symbiosis ever-since our first encounter in 2008, initially formed under the pressure of a highly demanding European project. Athanasios Karamalis, the gentle and kind giant, both from a human and a professional perspective. The past year, sharing an office with Tassilo and Athanasios, is already an unforgettable experience. Max Baust, for giving the initial boost to my post-EU-project research and helping me with my first MICCAI publication, while constantly wisecracking and making complex stuff seem so easy. Olivier Pauly, for his insights into Machine Learning (which lead to our project's second MICCAI publication) and for being the only other PhD I know that also lives part of his private life singing and shouting into a mic on-stage (we'll both become famous soon, I'm sure). It was really fun working so closely together - but more than that, I hope that the friendship that has developed over the years will last for many years to come.

Many many thanks also to our clinical partners, Prof. Dr. Kai Bötzel and Dr. Annika Plate, who had unbelievable patience during the initial years of our collaboration, until we got everything running. I really believe that in the end, we have accomplished something great together, and this is only because of the mutual respect and enthusiasm since the very beginning - also to you, thanks for four years of fun!

Also, I would like to thank the first generations of PhDs at CAMP, Wolfi, Marco, the three Martins, Jörg, Tobias, Ben, Darko, Hauke, Thomas etc. for building up the core of CAMP and forming a positive working environment together with Nassir for generations of PhDs to come.

In administration terms, Martin Horn, Martina Hilla, Jessica Weinig and the EU-project administrator ladies from the central administration were indispensable! There are so may others in - and outside of CAMP that I have probably forgotten...

Last, but not least, I would like to thank my parents - the best parents I could have wished for - for bringing me up onto this path, for their constant, love, support and belief in me. Thanks to my brothers Ebi and Ali and all my close friends, Tino, Lea, Chris, Robert/Katrin/Michi from ROAR, the Stusta kids... for keeping me sane throughout the hard times and for their amazing humor and joy for life that I have the pleasure of sharing every day. :)

Thanks to all! And now it's time to submit...!

Contents

1 Introduction

Parkinson's disease is a neuro-degenerative disease that has been known since the early 19^{th} century. So far, neither the causes for development of Parkinson's disease (PD) are fully known, nor a cure for the disease has been found. While the search for the possibility of a cure for PD is still ongoing, one branch of research is dedicated towards improving the diagnostic methods which allow for detecting PD, in particular at an early stage, since an early detection may help in treating the disease symptoms and decelerate its progression.

Neurological research of the past two decades has shown that signs of PD can be detected using transcranial ultrasound (TCUS), a non-invasive sonographic technique which allows for imaging of central brain regions involved in the control of movements. Tissue alterations in these regions associated with PD can be visualized as local bright speckle patches, so-called hyper-echogenicities, in the area of the substantia nigra (SN). These changes may be visible up to several years before typical motor symptoms of PD occur, i.e. several years before the typical time point for PD diagnosis today.

The research of this thesis is dedicated towards an extension of this sonographic technique from 2-dimensional (2D) to 3-dimensional (3D) imaging, and the development of techniques for automatic quantification of SN hyper-echogenicities. We present the first steps towards a system for computer-aided diagnosis and early detection of PD, which is robust and more objective than the current method in 2D, while still being applicable in clinical routine. At the end of this chapter, **in section 1.4, we give a detailed list of our main contributions**, towards achieving these goals. We also give a structural overview of this thesis and the main publications created during our research.

In this chapter, as an introduction, we first describe Parkinson's disease, i.e. its history, symptoms, treatment and state-of-the-art methods for diagnosis. Next, we explain the relevance of transcranial ultrasound (TCUS) for PD diagnosis and its potential for early detection of the disease. We recapitulate the neurological research of the past two decades on TCUS for PD diagnosis and summarize the diagnostic capability of this technique for PD, *in order to lay a solid foundation and medical argument for the work performed in this thesis*. We continue by describing issues with the TCUS technique by citing critical

remarks by other researchers in the medical community and how we hope to improve on the shortcomings of the TCUS technique by extending the method to 3D. Based on this, we enlist the contributions made in this thesis and give an outline for the overall structure of this document.

1.1 Medical Background - Parkinson's Disease

Transcranial ultrasound (TCUS) has been applied for Parkinson's disease (PD) diagnosis and early detection of the disease. Next to PD, TCUS has also been shown to have diagnostic value for several other neurological movement disorders. Since PD is one of the main clinical applications to which our contributions in this thesis can be applied, we will explain the most important facts about this disease in the following.

Frequency: PD is a disease mostly affecting elderly patients. Most patients present themselves with an onset of the disease between 55 to 66 years [102]. Approximately 1 million individuals in the US alone are affected. In the US, PD is the second most common neurodegenerative disease next to Alzheimer's disease [102].

History: Parkinson's disease (PD) is a neurodegenerative disorder of basal ganglia in the midbrain area. The disease is named after the English physician James Parkinson (1755 – 1824). He observed similarities across a group of patients showing motor disfuctions such as "involuntary tremulous motion, with lessened muscular power [...], a propensity to bend the trunk forward, and to pass from a walking to a running pace; the senses and intellects being uninjured". He described six cases, three of them observed by himself and three of them reported to him, in his seminal publication "An Essay of the Shaking Plasy" in the year 1817 [112]. Triggered by this publication, medical research observed many more cases and lead to a finer classification of the disease, with additional symptoms such as rigidity being classified and highlighted by Charcot and Vulpian in 1861 [97].

Parkinson subtypes and symptoms: Since its discovery, Parkinson's disease (PD) has been further investigated and a larger family of diseases have been identified, all of which are summarized under the term "Parkinsonism". There are four main subtypes of Parkinsonism, summarized by Jankovic [77]:

- *Idiopathic Parkinson's disease (IDP):* Accounting for around 90% of all cases of Parkinsonism, IDP is its most common form, and is hence also called primary subtype. Idiopathic means that the concrete reason for development of the disease is unknown.
- *Acquired or symptomatic Parkinson's disease:* Also called the secondary subtype of PD, acquired PD is caused as a side effect of e.g. medication intake, other neurological disorders or other illnesses. In contrast to idiopathic PD, the secondary PD subtype can be reversible, if the cause

of PD can be identified and if no permanent neurological damage has been caused so far.

- *Hereditary Parkinsonism:* In general, PD is of non-genetic origin, however, in 15% of cases, PD patients have a first-degree relative who also were diagnosed with the disease and specific types of genetic mutations have been shown to cause PD in around 5% of patients [98].

- *Atypical Parkinson's disease:* This subtype, also termed "Parkinson plus syndrome", describes cases of Parkinson-like symptoms with additional symptoms. Typical forms are multiple systems atrophy (MSA), progressive supranuclear palsy, corticobasal degeneration and dementia with Lewy bodies [133].

Concerning symptoms, there are early symptoms, which might indicate onset of the disease and actual disease symptoms. Early symptoms are difficult to detect and often have to be recollected by patients after onset of the main PD symptoms during anamnesis discussions with the treating physician. As Parkinson denotes [112]: "The first symptoms perceived are a slight sense of weakness, with a proneness to trembling in some particular part; sometimes in the head, but most commonly in one of the hands and arms."

After onset of the disease, the main symptoms of Parkinsonism are separated into motor and non-motor symptoms [77].

There are four main motor symptoms of PD which are often referred to as the "cardinal" symptoms of PD [77]. These are tremor at rest, rigidity, slowing of movement and postural instability. These symptoms are not always present in patients at onset of PD. For example, while the symptom tremor is the most common PD symptom, around 30% of individuals do not show this symptom in the beginning of the disease. It is hence important to note that it can be *difficult to detect PD at very early stages* of the disease, in particular at time points *before typical motor* symptoms occur [77].

Non-motor symptoms of PD include autonomic dysfunction, neuropsychiatric effects, sleep difficulties as well as sensory deficiencies such as loss of smell [77]. The most common neuropsychiatric effects include changes in mood, cognitive abilities and behavior. For example, around 50% of patients develop a depression during the progression of the disease [102].

Anatomical background and cause: The midbrain, also called mesencephalon, is an approximately 2x2x1cm sized region of the brain, which forms the upper part of the brainstem. It contains several conglomerates of nerve cells, so called nuclei, which are involved in visual, auditory, and motor functions of the human brain [2].

It is well known that a degeneration of cells in one type of midbrain nuclei, namely the Substantia Nigra (SN) (see figure 5.1), lies at the source of Parkinson's disease (PD) [168]. In PD patients, cells in the SN decay slowly throughout the progression of the disease. One of the functions of nigro-striatal cells in the SN is to produce dopamine, which is an important neurotransmitter involved in motor functions. The death of these nigro-striatal cells causes a

depletion of dopamine in other parts of the brain, thereby causing the cardinal motor symptoms of Parkinson's disease, i.e. slowed movement, tremor, and muscular rigidity.

The decay of SN cells is accompanied by a build-up of ferrite deposits. It is argued that these deposits cause slight hyper-echogenicities (i.e. local, bright speckle patches) in the SN area, if visualized with transcranial ultrasound (TCUS). *The main motivation of this thesis is the computer-aided detection and analysis of these hyper-echogenicities in 3D-TCUS.*

Pathophysiologically, PD is categorized as a synucleopathy, which means that a protein normally used for membrane stabilization (alpha-synuclein) aggregates for unknown reasons and forms pathological inclusions. These are known as 'Lewy bodies', named after their discoverer, Fritz Heinrich Lewy, in 1912. The mechanism by which cell death is triggered is unknown. Histological post-mortem analysis of SN in PD patients show neuronal loss and accumulation of Lewy bodies, identifying them as a key pathological feature of PD [111].

Treatment methods: Just as the exact mechanism of the disease is unknown, there is also no cure for PD available at present. However, there exist both non-surgical and surgical treatment methods which can partly reduce the symptoms.

Usually, therapy is commenced non-surgically through administration of drugs, mainly dopamine agonists and monoamine oxidase type B (MAO-B) inhibitors in the early stage of the disease and levodopa (L-DOPA) in the later stage. While the first two types of medication are not as effective in improving motor control as L-DOPA, the latter can lead to motor control complications induced by the drug itself [157].

In later stages of the PD, once control with medication achieves only unsatisfactory results, or if medication is inefficient early on, there is also the possibility for surgical treatment [158]. The most common form of surgical treatment of PD is Deep Brain Stimulation (DBS). In DBS, an electrode is implanted into basal ganglia around the midbrain area, such as the Subthalamic Nucleus in PD. The target region is stimulated with a stimulation device which is often referred to as a "brain pacemaker". Consequently, motor dysfunctions can be significantly reduced, sometimes to an extent making further administration of drugs unnecessary. Next to PD, a wide area of movement and also neuropsychiatric disorders can be treated, including dystonia [19] and depression [100]. Recently, patients with further disorders such as epilepsy, obsessive-compulsive disorder, substance abuse, Alzheimer's-type dementia and traumatic brain injury have undergone DBS surgery [134]. Side effects, however, can be caused by misplacement of the electrode and unwanted stimulation of adjacent brain areas, leading to e.g. visual impairment or further neuropsychiatric effects [81].

The relevance of early diagnosis of PD: Recent studies and meta-reviews have argued that it is beneficial to detect PD at an early stage and start treatment at time of diagnosis, since an early administration of certain drugs such

as levodopa [160] or the MAO-B inhibitor rasagiline [147] can decelerate the further progression of the disease and potentially delay motor symptoms, if PD was detected at a pre-motor stage. For example, Schapira and Obeso conclude in their review [135] that early administration of dopaminergic drugs and symptomatic treatment lead to an "early restoration of of basal ganglia physiology", which will "support compensatory events and delay the irreversible modification of circuitry that characterizes the clinical progression of PD. Such an effect will lead to lasting clinical benefit for the patient".

Diagnostic methods and usage of medical imaging methods: According to Brooks [31], the diagnostic process usually begins with patients presenting themselves with symptoms of tremor, stiffness or slowness, indicating possible PD. However, medical imaging can play an important role even in early stage of diagnosis (cf. figure 1.1).

Medical imaging at this early stage of diagnosis is limited to non-invasive methods such as MRI and transcranial US. As Brooks reports [31], MRI scans can help in excluding the possibility that structural lesions are causing PD-like symptoms. However, it is important to note that up to now, MRI is not useful as a method to detect pathological changes in the SN [31], e.g. ferric deposits, since scan protocols with sufficiently high diagnostic sensitivity and specificity still have to be identified [105].

Transcranial ultrasound can be used for detection of hyper-echogenicities, due to pato-physiological changes in the SN. If lesions can be excluded as the cause due to MRI and if transcranial US yields an abnormally large hyperechogenic SN area, further imaging methods can be applied. A diagnostic gold-standard method is provided with Single-photon emission computed tomography (SPECT) or Positron emission tomography (PET), yielding almost perfect diagnostic reliability, with sensitivity of 97% and specificity of 100% [161]. In SPECT and PET, the dopamine metabolism of the brain can be visualized with specific, gamma-emitting radioisotopes. PD can be diagnosed since the metabolism of affected patients differs from that of healthy subjects. However, cost, duration and invasiveness for SPECT scans are comparably high. Eventually, a definitive diagnostic result can only be given post-mortem, after "histological demonstration of intra-neuronal Lewy body inclusions in the substantia nigra compacta" [31].

Within the spectrum of available imaging methods for PD diagnosis, the *role of transcranial US* can be seen as a low-cost and non-invasive imaging method with comparably high diagnostic accuracy [23, 168, 169]. It is therefore useful as a strong indicator for PD diagnosis [22], even at an early stage [24], motivating its usage for screening of large populations to make possible an early treatment of PD, prior to motor symptom stage [31].

In the following chapter, we will explain the history of transcranial US imaging for PD diagnosis, its strengths and shortcomings and how we intend to contribute to this field with this thesis.

Summary - Parkinson's disease (PD) facts relevant to this thesis:

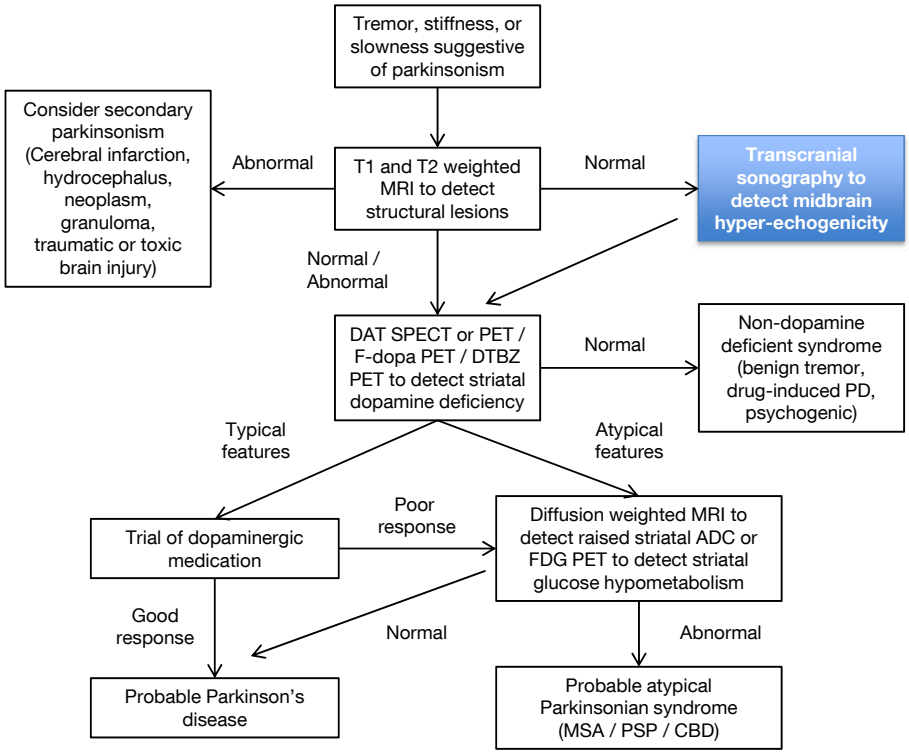

Figure 1.1: Medical imaging methods used for diagnosis of PD (image reproduced and adapted by courtesy of Professor David J Brooks MD [31], Head of the Centre for Neuroscience, Imperial College London.

- PD is relatively difficult to diagnose in early stage due to vague symptoms, which are often not detected by the patient themselves [77].
- Early detection of PD is relevant for containment and deceleration for nigro-striatal loss [24].
- Transcranial ultrasound (TCUS) is a small but important building block in the initial stages of PD diagnosis, among the wide range of applicable medical imaging methods as described by Brooks [31]. Compared to MRI, TCUS is complementary, since MRI is useful to exclude other lesions, but not yet useful for highly sensitive and specific detection of ferrite deposits in the SN [105].

1.2 Medical Motivation - Transcranial Ultrasound for Diagnosis of Neurological Movement Disorders

1.2.1 History of Transcranial Ultrasound for PD Diagnosis

In the past two decades, medical research mainly originating from research centers in Germany has investigated the usage of transcranial ultrasound (TCUS) for initial and differential diagnosis of neurological movement orders. First investigations in the mid-nineties discovered that pato-physiological changes in the substantia nigra (SN) of the midbrain can be observed in transcranial ultrasound (TCUS).

In a seminal work by Becker et al. [14], a study cohort of 60 subjects was scanned using the TCUS technique. Out of the 60 subjects, 30 subjects were previously diagnosed with Parkinson's disease (PD) and 30 subjects were used as a control group, showing no Parkinsonian symptoms. All subjects were scanned transcranially through the pre-auricular acoustic bone window, also called temporal bone window. The authors described anatomic structures visible in TCUS, if the ultrasound probe was tilted in axial direction while maintaining an axial cut-plane of the brain (see figure 1.2), paying particular attention to the "pontine and mesencephalic brainstem, the basal ganglia, the width of the ventricular system, and the supra-tentorial white matter" [14]. The echogenicity and echo-texture of these areas were compared between the PD group and the group of healthy controls. In 28 control subjects and 13 PD patients, the SN was undetectable by TCUS. Two non-PD controls and five PD patients showed a weakly hyperechogenic SN, while in 12 PD patients, the SN showed a distinctly enlarged hyper-echogenic tissue response in the area of SN. In terms of diagnostic accuracy, this amounted to a test with 40% sensitivity and 100% specificity. The authors also found a correlation between the amount of SN hyper-echogenicity and both severity and duration of PD ($p < 0.001$). Furthermore, the authors of this study identified a statistically significant enlargement of the mean widths of third ventricle in PD patients, compared to non-PD controls.

Triggered by these initial findings, other research groups investigated the value of TCUS scanning of the midbrain for diagnosis of neurological movement disorders such as PD. In the following, several groups and studies were able to confirm the results from [14]. Two notable groups having achieved excellent diagnostic value of the TCUS method include the group around Prof. Daniela Berg et al. [22] from University of Tübingen, Germany, and the group around Prof. Uwe Walter et al. [168] from University of Rostock, Germany.

In 2002, Berg et al. [23] associate the hyper-echogenic appearance of the SN in PD patients with *increased iron contents in the SN*, which is backed up by PET examinations in a group of alive patients as well as post-mortem histochemical analyses. In a paper from 2006, they hypothesize that SN hyper-echogenicity is a pre-motor symptom and can help in early detection of the disease and consequently be used as a basis for neuroprotection [20]. In 2010, an analysis of the intra- and inter-observer reproducibility was published [99], demonstrating

excellent agreement and correlation across four expert observers assessing SN hyper-echogenic area in 2D TCUS (cf. chapter 4.4).

While Berg et al. have systematically assessed the potential of TCUS for early detection of PD, the group around Walter et al. have focused on demonstrating the potential of TCUS for *differential diagnosis of PD* and also other neurological movement and neuropsychiatric disorders. In 2007, Walter et al. [168] investigated the usability of B-mode TCUS for differential diagnosis of PD and discriminating between Parkinsonism and idiopathic PD. The study cohort encompassed 138 patients with sporadic idiopathic PD, 21 with multiple-system atrophy (MSA-P) and 22 with progressive supranuclear palsy (PSP). In this cohort, 7 subjects had unsuitable bone windows, preventing TCUS scans (3.8%). Like Becker et al. [14, 13] and Berg et al. [23, 20], the authors found that the area of SN echogenicity was an important salient feature for distinguishing the groups. Different combination of TCUS features were able to distinguish atypical PD (MSA-P or PSP) from idiopathic PD with sensitivities of up to 84% and specificities consistently above 97%. Using only the SN hyper-echogenic area as a feature for distinguishing MSA-P from PD even resulted in a sensitivity of 90% and specificity of 98%.

Based on observations made during experiments with TCUS and PD diagnosis, reference groups formed the hypothesis that SN hyper-echogenicity emerges *notably long before a detectable onset of typical PD motor symptoms*, such as tremor or rigidity. Thus, SN hyper-echogenicity can serve as a disease marker and potential factor in early detection of PD, making TCUS a potential tool for early PD diagnosis in a cheap, non-invasive and widely available manner.

Very recently, this hypothesis was confirmed in a large, multi-center study, in which three clinical research centers combined their efforts over a duration of 37 months, resulting in the study of Berg et al. from 2011, [24]. This study tested 1847 individuals over 50 years age without any evidence for PD or any other neurodegenerative disease at baseline, i.e. at the beginning of the study. A final resulting group of 1535 subjects could undergo reassessment after 37 months. Among this group, 11 cases of incident PD were diagnosed in the follow-up period. In participants with enlarged hyper-echogenic SN at baseline, relative risk for incident PD was 17.4 times higher than for normo-echogenic participants. The study authors reported that elderly patients with enlarged SN hyper-echogenicity had a *"17 times higher and thus highly increased risk to develop incident PD"* [24]. They concluded that TCUS may be a promising primary screening method to define a risk population for imminent PD. Consequently, the hope of medical researchers is to be able to use TC-US as a screening method for PD in future, especially since as mentioned, the degeneration of SN cells *cannot* yet be visualized by Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) [105, 31].

1.2.2 Transcranial Ultrasound for Diagnosis of Other Neurological Disorders

With growing experience in the TCUS technique, there were also other pathological changes identified which are visible in TCUS and which relate to other

disorders than PD. Since then, TCUS has been shown to be a non-invasive, low-cost and fast method for identifying several neurological disorders-related pathological changes within the midbrain and its surroundings. There has been an increasing number of diseases that have shown signal alterations detectable by this method, including atypical Parkinson syndromes [16] [170] [169], depression [15] [13], dystonia [108] and multiple sclerosis [70] [171].

In their review on TCUS technique and applications, Skoloudik and Walter [150] summarize all known TCUS abnormalities, related diseases and discriminated conditions in a table, which we reproduce in table 1.1.

TCS abnormality	Related condition	Discriminated condition
SN hyper-echogenicity	Parkinson's disease	Atypical parkinsonian syndrome
	Corticobasal degeneration	Progressive supranuclear palsy
	Dementia with Lewy bodies	Alzheimer dementia
SN hypo-echogenicity	Restless legs syndrome	
BR hypo-echogenicity	Unipolar depression	Bipolar affective disorders
LN hyper-echogenicity	Idiopathic dystonia	Psychogenic dystonia
	Atypical parkinsonian syndrome	Parkinson's disease
	Wilson's disease	Parkinson's disease
CN hyper-echogenicity	Huntington's disease	Other choreatic disorders?

Table 1.1: Characteristic echogenicity changes of deep brain structures in PD and other neurological movement disorders (SN = substantia nigra, LN = lenticular nucleus, BR = brainstem raphe, CN = caudate nucleus) (from Skoloudik and Walter [150]).

1.2.3 Issues and Critique at the Technique

Given the high diagnostic value and good study results of the TCUS technique for neuropsychiatric and neurological movement disorders, TCUS has established itself as a diagnostic tool both in research as well as in clinical routine among several clinical groups. However, despite the promising properties of TCUS, there are also downsides and critical assessments in the medical community.

One issue of the TCUS method recurrently criticized is its dependency upon the examiner's experience, which is accompanied by a high subjectivity leading to a limited inter- and intra-observer reliability [149] [162]. A recent study which compared diagnostic reliability given different observers and different ultrasound systems showed divergent results and therefore concluded that the TCUS technique is *not yet ready for screening large populations* [162]. This meta-analysis found a broad range of sensitivity varying from 48-100% as compared to the final clinical diagnosis.

Another unresolved issue of this method is the dependency upon an adequate temporal acoustic bone window. Apparently, this aspect also causes inconclusive results in around 13% of cases (average of 35 studies, analyzed in a meta-review by Vlaar et al., [163]). It is important to note that the acoustic bone window quality is indeed a limiting factor of the TCUS technique. While there has been work investigating the compensation of phase aberration effects

occurring due to sound waves transversing the bone layer [75], there is currently no way to overcome the fundamental physical limitations of ultrasound penetrating thick bone layers and still achieve B-mode imaging. One possible way to still obtain ultrasound echoes through thicker skull regions is to transmit waves with very low frequency on the order of 0.5-1MHz, e.g. for trans-skull fluid detection [152] or intra-operative brain shift monitoring using shear mode transcranial ultrasound [176]. However, up to now, the signal response in their system is restricted to A-mode ultrasound [176], i.e. a temporally varying echo response on a single scanline (see chapter 2 for more details on ultrasound physics and modalities).

In summary, even if performed by expert sonographers, TCUS cannot be applied in approximately 10% of people, due to insufficient bone windows, as observed in the 1847-person study of Berg et al. [24]. Since no B-mode images can be obtained in 2D, the same failure rate applies to the 3D methods proposed in this thesis and can be seen as a general limitation of the TCUS technique.

While not a limitation, there is one further open question with the TCUS technique. It is not fully known how and why the hyper-echogenicities in the substantia nigra (SN) occur. As mentioned before, Parkinson's disease (PD) is accompanied by a build-up of ferrite deposits in the SN, although it is not known whether this is the cause or an effect of the disease. It is hypothesized but not yet confirmed that the pathological substrate of the hyper-echogenic abnormalities in PD are probably iron deposits [23], causing micro-scattering of ultrasound waves and the formation of hyper-echogenic speckle patterns.

1.3 Transcranial Ultrasound Specifics

In the last sections, we summarized the potential of TCUS for diagnosis of neurological movement disorders, but we also reported critical remarks. Similar to other application areas of medical ultrasound, TCUS is a difficult technique to master and requires long-term expertise until it can be properly applied for diagnostic purposes [161, 162].

In this section, we will explain the workflow and the most important parameters of transcranial ultrasound (TCUS) imaging and describe the advantages of this imaging modality which also make it attractive for screening of large populations. We will then continue by explaining image properties, artifacts and distortions which make TCUS overall difficult to master. In chapter 2, we will explain the physical backgrounds of TCUS imaging and medical ultrasound in general, as well as where these artifacts come from. Before that, at the end of this introduction, we will explain the contributions of this thesis and how we plan to overcome the shortcomings of TCUS with the methods proposed in this thesis.

Technique and Workflow: An international consensus guideline has been formed about the workflow of the TCUS examination, e.g. concerning patient positioning, transducer placement and imaging parameters. A concise and comprehensive explanation of the clinical workflow of TCUS examinations is given by Skoloudik and Walter in [150].

The patient is placed in supine position, i.e. lying with his back on a bed or in a leaning chair, with the physician at the head of the examination table. The transcranial ultrasound transducer is placed at the pre-auricular bone window, in posterior or middle position of the window, as seen in figure 1.2A. Pressing the TCUS transducer firmly to the bone window, it is first aligned to the orbitomeatal plane[1]. Then, the scan-plane is tilted in a sweeping motion to reach standardized axial scanning planes, as depicted in figure 1.2B. The solid line depicts the standard orbitomeatal plane for mesencephalon examination, in which "substantia nigra, red nucleus and brainstem raphe can be assessed" [150]. The dashed line depicts the axial scanning plane at thalamic level, at which the "third ventricle, frontal horns, thalamus, lenticular and caudate nucleus are assessed" [150].

Upon identification of the optimal scan-plane, the image is frozen and the measurement tools from the ultrasound machine are used to manually outline the relevant anatomic features. For example, for PD diagnosis, these are so-called SN hyper-echogenicities (SNE), i.e. irregularly bright speckle patches within the area of the SN. For diagnostic analysis, the size of the cross-sectioned 2D area of SNEs is measured and compared against a reference value, while relative brightness is not considered in clinical practice, the SNEs are manually outlined and the US machine measures the area in units of $[cm^2]$. An US-machine specific threshold for this area (e.g. above $0.25mm^2$ for SNE in PD diagnosis, [168]) then determines the diagnostic outcome.

[1] the orbitomeatal can be roughly described as the plane spanning between the eyes and the two ear canals

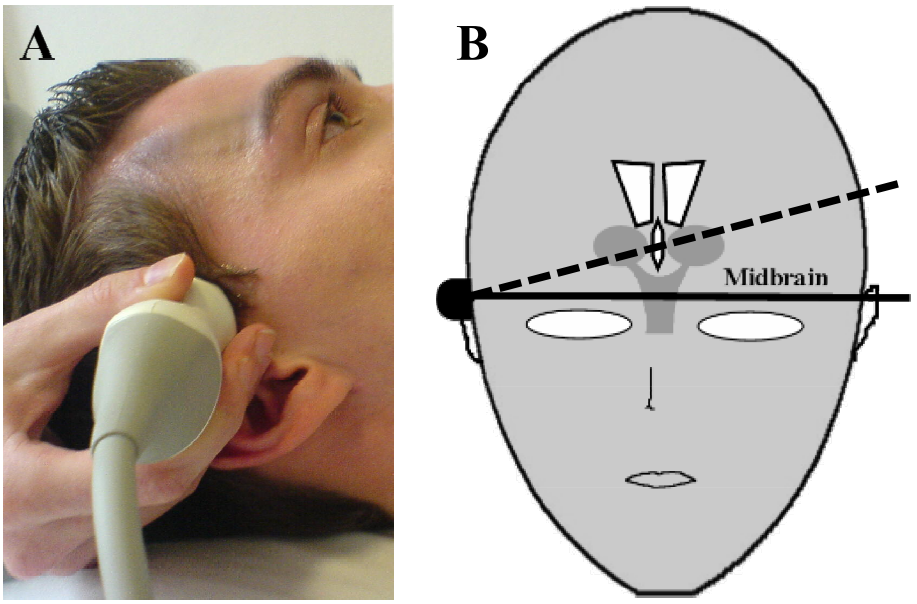

Figure 1.2: (A) Patient positioning and transducer location at pre-auricular bone window and (B) transcranial probe scan-planes for examination of midbrain, various basal ganglia and ventricular system, as described in [150] (images reproduced and adapted by courtesy of Prof. Dr. med. Uwe Walter, Associate Clinic Director, Klinik für Neurologie und Poliklinik, Universitätsklinikum Rostock, Germany, email for approval received 07. August 2012).

Imaging parameters: In order to be able to non-invasively penetrate the skull bone using ultrasound, the wave frequency with which the brain is being insonified has to be kept low. Operating frequencies of TCUS typically lie in the range of 2-5 MHz [21]. For better orientation during the 2D scan, it is helpful to see the opposite cranial wall in the B-mode image. This requires a large scan-depth of 14-16cm, which is reduced if needed [150]. Time-gain compensation (TGC) has to be set such that the deep scan depth is compensated, e.g. a linear ramp [168]. Dynamic range is set to 45-50 dB [103, 124, 150]. It should be noted that these settings vary depending on the ultrasound machine. While several recommendations for settings on wide-spread machines such as the Siemens Acuson Antares (Siemens AG, Erlangen, Germany) have been described before [168], it is advised to establish new optimal settings for a previously unused machine.

Scanning at low frequencies of 2-5 MHz has the advantage that sound can penetrate deep into the human brain, up until reaching the opposite cranial wall at a distance of roughly 14cm and giving a comparably large field-of-view given the phased-array geometry of transcranial transducers. Depending on the quality of the pre-auricular bone window, TCUS allows for a clear delineation of the midbrain and other anatomical structures, such as the opposite cranial wall, as visible in figure 1.3.

Shortcomings and challenges: Figure 1.3 also shows several challenges and shortcomings of TCUS. Depending on the physiology of the pre-auricular bone window, image contrast in TCUS can be low and anatomic boundaries can vanish (cf. figure 1.3, bottom left), up to a degree that makes the image unusable (cf. figure 1.3, bottom right).

Due to the low scanning frequency, image quality is relatively low and decreases further with increasing penetration depth. The large wavelength at frequencies of 2-4 MHz results in low axial and lateral image resolution (see section 2.1 for reasons and a general explanation of US image physics). For example, using an Elegra sonographic scanner (Siemens, Erlangen, Germany) and a scanning frequency of 2.5 MHz, axial image resolution lies at approximately 0.7 mm while lateral resolution lies at about 3mm, as reported in [124]. Scanning through the skull bone leads to significant absorption of sound energy, leading to low image contrast and signal-to-noise ratio (SNR). Furthermore, the large wavelength results in large and inconsistent speckle patterns which reduce the signal-to-noise ratio (SNR) further in the images. Scanning through skull bone also introduces non-linear and unknown distortions, especially at non-orthogonal probe angles, due to refraction phenomena. Also, common medical ultrasound machines make wrong assumptions about the speed of sound in brain and particularly skull bone tissue, resulting in de-focusing phenomena and aberrations of ultrasound wave phases [75].

The low contrast, SNR, missing anatomic boundaries and non-linear distortions make TCUS a *challenging image modality*, which partly explains why physicians require plenty of experience in order to be able to reach high diagnostic accuracy using 2D TCUS images [99, 163].

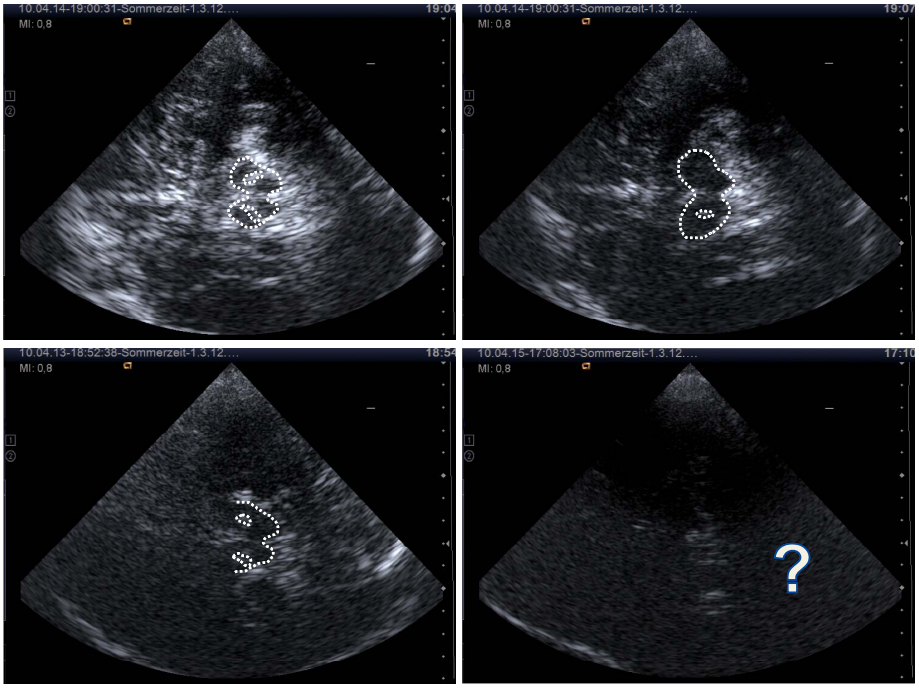

Figure 1.3: Four examples of transcranial ultrasound (TCUS) images in 2D. The midbrain and substantia-nigra hyper-echogenicities within the midbrain are highlighted with a white dashed outline. Top left: the example shows a TCUS image quality given an optimal bone window, leading to comparably high contrast and complete anatomic outlines of midbrain and SN. Top right: the image was taken through a medium-quality bone window with lower contrast and less visible anatomy surrounding the midbrain. Bottom left: the image shows a low-quality situation, with partly missing midbrain boundaries, low contrast, few visible anatomic features and low contrast of the opposite cranial wall. Bottom right: in about 10% of subjects, the bone window does not allow sufficient sound energy to penetrate the skull [24]. This leads to un-interpretable images, both in 2D and 3D.

1.4 Thesis Contributions and Structure

As explained, the TCUS in 2D is a very promising technique with high potential and possibly far-reaching consequences for detection and diagnosis of Parkinson's disease and other neurological movement disorders, in particular *at an early stage*. However, we have also described the critiques of this technique and the challenges of this image modality.

In this thesis, we propose several methods for making the TCUS technique easier to use and more objective, while expanding the research possibilities in terms of diagnostic value due to the extension of the technique into 3D.

In the following, we will explain the contributions of this thesis in detail, where each contribution will be laid out by dedicated chapters:

Chapter 3 - System design and first acquisitions of TCUS data in 3D: We propose, for the first time, to extend the TCUS examination method to 3D by using a 3D Freehand Ultrasound approach. Through an additional reference tracking target attached to the forehead, we acquire data bi-laterally and reconstruct it into a joint volume, which partly allows for compensating low-quality bone windows. We record an unprecedented 3D-TCUS dataset with a significant cohort-size of 11 diagnosed PD patients and 11 healthy controls for comparison between the groups. Using manual expert segmentations, we perform the first quantitative analyses of 3D SN hyper-echogenicities and report first volumetric measures of anatomy. The system design and initial data analysis were incrementally presented at multiple medical conferences [117, 120, 118] and eventually awarded by the German Parkinson Association in 2011 [119].

Chapter 4 - First multi-observer study and first steps towards computer-aided diagnosis: Using two different segmentations of the 22-subject dataset by an expert and novice observer, we perform the first multi-rater study based on volumetric measures of anatomy. We report the first values for correlation between volume measures and compare them to 2D, given a first indication that the objectivity of TCUS is increased by using analyses in 3D. Using decision-making and machine-learning methods, we demonstrate that the patient- and control-groups can be separated with sensitivities and specificities of up to 91% and 72%, respectively. The results have been published in the Journal for Ultrasound in Medicine and Biology [121].

Chapter 5 - (Semi-)automatic midbrain segmentation: Manual segmentation of target anatomies in 3D ultrasound volumes is tedious and would take too much time if executed in clinical routine. We therefore propose to perform the segmentation of midbrain and SN hyper-echogenicities in a computer-assisted manner. In chapter 5, we propose a computer-aided segmentation method for the midbrain, since this facilitates the following step of SN detection. Previous work in literature is restricted to 2D midbrain segmentation and is often performed in a manual manner. In contrast to that, we present a (semi-)automatic approach for volumetric segmentation of the midbrain in

3D ultrasound volumes, based on an explicit active surface method, which is extended using a midbrain shape prior and localized region-based cost function for robust and accurate performance. The approach was published at the International Conference for Medical Image Computing and Computer Aided Intervention (MICCAI) 2011 [2].

Chapter 6 - Automatic detection of Substantia Nigra Echogenicities: Once the midbrain outline has been identified using manual or computer-aided segmentation, the inside voxels have to be classified to identify SN echogenicity regions, which eventually yields the diagnostically valuable information. While previous work is again restricted to 2D B-mode images and ranges from morphological operations to supervised classification of intensity features, we propose a probabilistic approach based on Random Forest classification. Using illumination-invariant local intensity descriptors and a novel prior incorporating the spatial location of the SN within the midbrain, we perform a fuzzy classification of midbrain voxels, leading to a probabilistic posterior map of SN voxel probabilities. The approach was published at MICCAI 2012 [113].

Chapter 7 - Towards improved reconstruction of bi-lateral TCUS volumes: When acquiring transcranial ultrasound images, scanning through the skull bone introduces non-linear distortions, spatial mis-registration and phase aberrations. Slight mis-calibration of the 3D Freehand Ultraosund system leads to additional linear registration errors between the bilateral volumes. In this chapter, we propose a first step towards a compensatory reconstruction technique, which uses a sequence of inter-dependent segmentation, registration and reconstruction steps (JSR2) to combine images taken from the left and right bone window into one joint volume. We present first results on simulated midbrain 3DUS data and on several cases of the 22-subject dataset.

2

Basics of Ultrasound Imaging

In order to better understand the image formation of transcranial ultrasound (TCUS) images and medical ultrasound in general, this section is dedicated to explaining the basic *physical properties of ultrasound*. We will explain the fundamentals of wave formation and its propagation in human tissue, including phenomena like transmission, reflection and scattering. We will furthermore explain how these phenomena can be generated and measured using ultrasound transducers in order to create the grey-valued images which are usually used in clinical practice, including 2D images as well as 3D US image volumes. At several points, we will draw conclusions related to TCUS imaging, e.g. by explaining the physical background for typical TCUS image artifacts and challenges described in the introduction.

2.1 Basic Ultrasound Physics

We begin by explaining the basic principles of wave generation and propagation of wavefronts emanating from point-sources. Next, we explain the principles of interaction of a single ultrasound beam with tissue. This lays the foundation for ultrasound imaging and provides basic explanations for tissue appearance in actual ultrasound images (see section 2.1.2). At the end of this section, we revisit our medical problem and explain implications of learned physical US principles for TCUS imaging.

2.1.1 Wave generation and propagation

Ultrasound is a type of sound wave which propagates through human tissue as its medium. The wave is generated by mechanical excitation through a sound inducer. In medical ultrasound, these inducers are typically piezo-electric elements. The front face of each piezo-electric element serves as a membrane

which can generate oscillating high-frequency excitations in the MHz range. The excitation is created spatially close to human tissue and by using acoustic coupling measures such as coupling gel, the excitation is transferred into human tissue through the piezo-electric membrane, which pushes the adjacent tissue. This creates high-frequency regions of compression and rarefaction of the molecules in the tissue [67]. Through neighbor-to-neighbor interactions, these compression and rarefaction regions are propagated through the tissue, creating the ultrasonic waves which lie at the fundament of medical ultrasound imaging.

Just like electromagnetic waves, ultrasound waves have basic properties, such as speed of sound c, frequency f and wavelength λ. These three wave parameters have the following important, linear dependency [67]:

$$c = f \cdot \lambda \tag{2.1}$$

From equation 2.1, we can see that the speed of sound with which the wavefront travels through human tissue depends linearly on frequency and wavelength of the excitation. Since the speed of the soundwave is approximately constant in human tissue, increasing the frequency leads to a reciprocal decrease of the wavelength.

Ultrasound is defined as mechanical waves with frequencies higher than the range that is audible to humans, i.e. with frequencies exceeding 20 kHz. However, the frequencies of medical ultrasound lie much higher, typically in the range of 1-40 MHz. The average speed of ultrasound in human tissue assumed by modern medical ultrasound scanners lies at 1540 m/s. Based on the above explained *inverse-dependency* between ultrasound frequency and wavelength, we can calculate that medical ultrasound wavelengths lie between 1.5mm and 0.038mm.

It is important to note that medical ultrasound devices assume an average and constant speed of 1540 m/s for human tissue. However, the actual speed of sound depends on the tissue type. Table 2.1 enumerates the speed of sound in different types of human tissue.

Axial, lateral and temporal resolution: In the following, we will explain three types of resolution important in an ultrasound imaging system, namely longitudinal, lateral and temporal resolution.

The *longitudinal resolution* (also called axial, radial or depth resolution) describes the resolution along the ultrasound beam, i.e. in direction in which the wave is traveling. It describes the smallest possible distance between two objects so that the ultrasound system can still display them as separate objects. This resolution is determined by two factors, the wavelength λ and the spatial pulse length (SPL) [67]. The SPL describes the length of the pulse which is modulated onto the carrier wave with a certain wavelength λ (cf. figure 2.1) and often is a multitude of λ, depending on the number of wave cycles n which the ultrasound system uses to form the pulse. The longitudinal resolution r can be calculated from these two factors as:

$$r = \frac{SPL}{2} = \frac{n \cdot \lambda}{2} = \frac{n \cdot c}{2 \cdot f} \tag{2.2}$$

Material	Density (kg/m3)	Velocity (m/s)	Acoustic impedance (kg/m2/s)
Air	1.2	330	0.0004
Water (20°C)	1000	1480	1.48
Liver	1060	1550	1.64
Muscle	1080	1580	1.7
Fat	952	1459	1.38
Brain	994	1560	1.55
Kidney	1038	1560	1.62
Spleen	1045	1570	1.64
Blood	1057	1575	1.62
Bone	1912	4080	7.8
Lung	400	650	0.26
Soft tissue average (e.g. abdomen)	**1060**	**1540**	**1.63**

Table 2.1: Speed of sound in different types of soft tissue in the human body (from [67]).

As this formula states, the minimum resolvable distance between two objects is directly dependent on the wavelength and indirectly dependent on the system frequency, i.e. higher frequencies allow for resolving smaller structures and thus increase logitudinal resolution.

The *lateral resolution* (also called angular, azimuthal or transverse resolution) describes the ability to "distinguish, as separate entities, two objects adjacent to each other oriented perpendicular to the beam axis" and is thus "a major factor in the quality of diagnostic ultrasound imaging" [67]. Lateral resolution is determined by the beamwidth of the imaging system, which for its part is determined by factors such as ceramic geometry, frequency, focusing and distance of the transducer [9]. For focusing, several single-element piezo-crystals are put in a geometric arrangement and excited with a carefully chosen set of time delays for each element, reducing side lobes and focusing individual wavefronts into a spot of high sound intensity at distance F in front of the transducer face (see also section 2.2 and figure 2.2). The formula for beam width w of a single transmitter piston can be given as [67]:

$$w = \frac{1.4\lambda F}{2a} = \frac{1.4Fc}{2af}, \tag{2.3}$$

where $2a$ stands for the aperture of the transducer's front face and F stands for the focal length of the transducer.

The *temporal resolution* of the system describes the temporal update rate with which new images (single scanlines, 2D images or even 3D volumes) can be generated. The temporal resolution depends mainly on line density and penetration depth [51, 67], since the machine has to wait until the furthest desired tissue has been insonified and the echo has traveled back to the transducer. For example, at a penetration depth of 15cm and an assumed speed

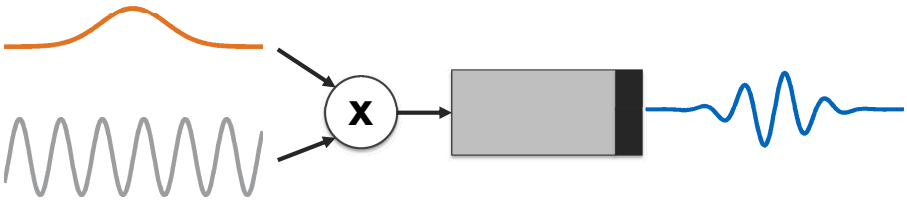

Figure 2.1: Modulation of an excitation pulse (orange curve) onto a carrier wave (grey curve) with wavelength λ. The resulting modulated pulse (blue curve) is used for excitation of a longitudinal ultrasonic wave insonifying the tissue.

of sound of 1540 m/s, the travel times for transmit and receive of a single scanline amount to approximately 0.195ms. A typical 2D ultrasound image is made up of several scanlines. Assuming 128 scanlines, the time necessary to acquire the information for a single image takes roughly 25ms, resulting in a maximum possible update rate of 40Hz.

2.1.2 Interaction with Human Tissue

The interactions between the ultrasonic wave and human tissue cause similar wave behavior as observed in light, including reflection, transmission, refraction, scattering, diffraction, divergence, interference and absorption [67].

Different types of tissue in the human body have different acoustic impedances Z, which can be calculated using tissue density ρ and speed of sound c (see table 2.1 for examples of speeds of sounds in different types of human tissue), i.e.:

$$Z = \rho \cdot c \tag{2.4}$$

The so-called *impedance mismatch*, i.e. the difference in acoustic impedances between adjacent tissue layers, causes *reflection*, the most important physical effect in medical ultrasound imaging. For two tissue types with impedances Z_1 and Z_2 (with tissues Z_1 and Z_2 lying proximal and distal to the transducer respectively) and for perpendicular incidence of the sound beam onto the tissue boundary, we can calculate the reflection coefficient α_r, which describes the percental sound energy that is reflected at the tissue boundary:

$$\alpha_r = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2 \tag{2.5}$$

The amount *transmitted* at the tissue boundary and allowed to travel further into tissue can be calculated as:

$$\alpha_t = 1 - \alpha_r = \frac{4Z_1Z_2}{(Z_1 + Z_2)^2} \tag{2.6}$$

For specular reflection, i.e. for perpendicular incidence of the sound beam onto the tissue boundary, these equations for transmission and reflection are independent of the frequency [67].

Another important physical phenomenon in medical ultrasound is *scattering*. Scattering occurs when the wavefront hits tiny reflectors, whose size is on the order of the wavelength or smaller. Each of these small scatterers cause a reflection of the soundwave as a point-source. The overlay of such a myriad of acoustic point-sources cause a complex pattern of constructive and destructive wave interferences [67], resulting in an overall brighter appearance for a large number of micro-scatterers. For example, a high amount of these reflectors are present in soft tissue, causing a so-called hyper-echogenic (brighter) appearance in US images, while fluid-filled cavities such as cysts contain less micro-scatterers and thus appear in a so-called hypo-echogenic (darker) manner.

Depending on the speed of sound between two adjacent tissue types, another type of sound wave interaction occurs, namely *refraction*. If a sound wave front hits the tissue boundary in a non-perpendicular angle, the sound wave does not travel further in a straight line after the tissue boundary, but experiences an angular deviation from its original path. The angular deviation is described by "Snell's law" [67]:

$$\frac{c_i}{c_t} = \frac{\sin\theta_i}{\sin\theta_i}, \tag{2.7}$$

where θ_i is the incident angle, θ_t is the transmitted angle, c_i is the speed of sound in the incident tissue and c_t is the speed of sound in the transmitted tissue.

In contrast to reflection and refraction, where ultrasound energy was returned to the transducer, the process of *absorption* causes a dissipation of energy into tissue and a conversion into other energy forms, mostly heat [67]. The amount of absorption is determined by the speed of sound, the viscosity and the relaxation time of the tissue. Scattering also causes a loss of energy and the combined effects of scattering and absorption result in the so-called attenuation of ultrasonic beam energy. The reduction in acoustic energy due to the combined effects of absorption and scattering follows an exponential function [67]:

$$p_0 = p_{\max} \exp\left(-az\right), \tag{2.8}$$

where p_0 is the peak pressure amplitude of the beam at depth z, $p_{\max}$ the initial peak pressure amplitue of the beam, z the distance of the wavefront from the transducer face and a the attenuation coefficient. The coefficient $a = a_s + \alpha$ is a sum of of the scattering coefficient a_s and the absorption coefficient α. The unit for these coefficients is called *neper* [Np].

For example, brain tissue causes an attenuation of $a = 0.098\text{Np/cm}$, while skull bone has a more than 20 times higher value of $a = 2.3\text{Np/cm}$, both at a frequency of 1 MHz [67]. In terms of decibel (dB), the attenuation in brain tissue at 1 MHz amounts to a energy reduction of 0.85 dB per cm, while in skull bone, the reduction is as high as 20 dB per cm. The half-value layer (HVL) or half-value thickness of material is the thickness of tissue that will reduce the sound intensity to half its original value. For a scanning frequency of 2 MHz, which is common to TCUS imaging, the HVL of brain tissue is 2cm [67].

2.1.3 Implications for Transcranial Ultrasound

In summary, there are several important physical properties of wave propagation which are relevant to the topic of this thesis:

- Visual echoes in brain ultrasound images are caused by wave reflections at boundaries between brain tissue regions with different acoustic impedances.
- The speeds of sound in skull bone and brain tissue are different from the average speed of sound in human tissue assumed by medical ultrasound devices. This causes aberrations of the ultrasound wave phase [75] and non-linear distortion of the image. Also, the ultrasound beam is defocused, causing a decrease in lateral image resolution and an overall deterioration of image quality. Furthermore, this causes a mis-assessment of distance between ultrasound transducer and tissue boundaries, also referred to as longitudinal mis-registration [67].
- Skull bone tissue is a highly attenuating material and the attenuation increases linearly with frequency. Thus, low-frequency ultrasound on the order of 2-5 MHz has to be used to penetrate the bone layer [150], which in return reduces longitudinal tissue resolution and leads to large speckle patterns.
- Microscopic objects with sizes on the order of ultrasound wavelength or smaller cause micro-scattering. Such areas appear brighter in the ultrasound image and are hence called hyper-echogenic. As hypothesized by reference groups in TCUS [23, 168, 99], ferrite deposits in the area of decayed SN cells cause micro-scattering and are responsible for the hyper-echogenic appearance of the SN in Parkinson disease patients. However, so far it has not been possible to confirm this hypothesis.
- According to Snell's law, sound waves hitting tissue boundaries at non-perpendicular angles are non-linearly refracted. At the interface of skull bone and brain, this refraction is important for the most lateral US scanlines in the 2D image, i.e. the further away from the midbrain, the more the sound beams got refracted, leading to increased de-focusing and distortion in the lateral image regions. In TCUS, this is visible as notably decreasing image quality towards the lateral edges of the US fan (see figure 1.3).

2.2 Beamforming and Transducer Geometries

Medical ultrasound systems feature a range of transducers types which can be applied for different application areas. Transducers are made up of several pistons which create acoustic pulses that are transmitted into tissue. Transmission occurs only if sufficient acoustic coupling is provided, which is why transducer housings provide acoustic bonding and ultrasonic coupling gel is provided to eliminate the air gap between the transducer surface and the patient's skin. There are various possibilities for geometric arrangement of piezo-elements in transducers. Three common geometries are linear, curvilinear and phased-array transducers.

Beamforming: In order to create beams with low lateral extent and thus high lateral resolution, several transducer elements at a time are used to generate an electronically focused beam. A sub-group of elements from the overall transducer arrays fire acoustic pulses with a previously computed sequence of individual time delays (see figures 2.2A and 2.3). The time delays are computed such that the individual wavefronts create destructive interference in all angular directions from $\theta = [-90°...90°]$ except in axial direction, i.e. for $\theta = 0$, where constructive interference ideally should lead to a laterally narrow acoustic pulse with high energy. This basic principle of electronic beam-focusing is called delay-and-sum beamforming [83], i.e. in a sub-array with M piezo-elements, the output signal $z(t)$ can be defined as an amplitude-weighted and time-delayed sum of the M individual elements' acoustic impulse responses $y_m(t)$:

$$z(t) \equiv \sum_{m=0}^{M-1} w_m y_m(t - \Delta t_m), \tag{2.9}$$

where the amplitude weights w_m are called the array's shading and the set of time delays Δt_m is called the array pattern [83]. Array shading and pattern are optimized such that the beam's shape becomes highly focused and side lobes become suppressed as much as possible. Side lobes are un-desired fields of high acoustic energy in non-axial angular direction, i.e. $\theta \neq 0°$. Since side-lobes can be considered as weakened beams into other, un-desired angular directions, they create tissue responses from areas which are un-desired for the current beam. This leads to image clutter and angular mis-registrations, which is why ultrasound transducer and systems are designed to suppress side lobes as much as possible.

By changing the time delays of the array's sub-group of elements, the beam can not only be focused, but the wavefront can also be directed into a certain direction (see figure 2.2B). This process is called electronic beam-steering [83]. Through both focusing and beam-steering, a transducer array can be utilized to scan the tissue with acoustic rays in spatially "discrete" directions which are called scan-lines.

Proper array shading and array patterns are usually implemented in the ultrasound machine and are only manually variable if a research interface is provided (e.g. as in the Sonix RP ultrasound machine with research SDK, Ultrasonix, Toronto, Canada). Since the exact theory and optimization of

proper beam-forming and beam-steering techniques are out of scope for this thesis, we refer the interested reader to [83] and [80].

Transducer geometries: As mentioned, piezo-elements in the transducer are commonly arranged in a *linear* or *curvilinear* arrangement, as seen in figure 2.3.

Linear arrays can provide good temporal and spatial resolution, but the transducer's ROI is limited by the width of the array. Hence, if a wide ROI is desired, the large width of the transducer's footprint makes it difficult to properly maintain contact to the patient's skin surface [67].

Curvilinear arrays produce large fan-shaped sector images since the transducer elements are arranged along an arc. Similar to linear arrays, groups of elements are fired together in order to generate focused acoustic pulses with high-energy wave-fronts. Scan-lines are oriented in perpendicular direction from the curved transducer surface. The focus qualities at the edges of the fan are similarly high as in the image center, but due to the angular orientation of scan-lines, the "gaps" between rays increase with higher distances from the transducer surface.

This disadvantage is also present in phased-array transducers. As described, the fan-shaped image geometry is generated artificially through electronic beam-steering. Analogously, the lateral image resolution decreases radially from the transducer surface, and the "gaps" between acoustic rays have to be filled with polar interpolation in the so-called scan conversion process (see chapter 2.3). In contrast to curvilinear arrays, however, focusing properties for scan-lines at the fan edges deteriorate compared to rays around the centerline [67, 83].

With regard to transcranial imaging and this thesis, it should be noted that transcranial ultrasound is mostly performed with *phased-array* transducers, due to their small footprint and the possibility to acquire images through a small window with a relatively large ROI. However, as Hedrick notes, imaging with a linear phased-array at low frequencies increases the number and intensities of side lobes for each element [67]. Additionally, the regular spacing of elements in the array causes so-called grating lobes, which occur due to summation of low-intensity side-lobes from the individual array elements. These secondary lobes can contain significant energy and lead to clutter, reduced image contrast and spatial mis-registrations, in particular at highly reflective interfaces (e.g. brain tissue to contra-lateral skull-bone) [67]. Additionally, side- and grating-lobe suppression becomes increasingly difficult for large steering angles [67, 83]. These physical phenomena provide further explanation for the low image quality of TCUS.

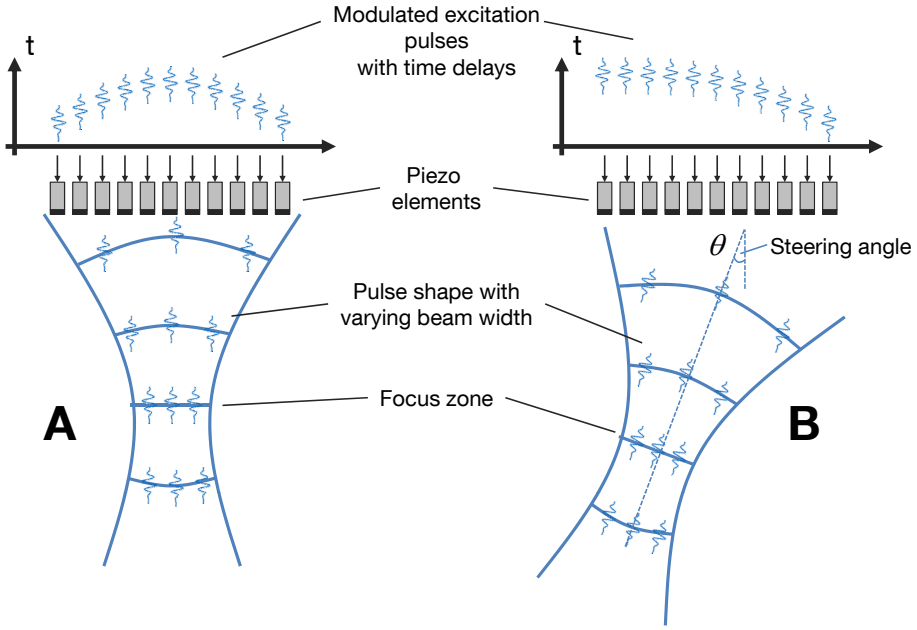

Figure 2.2: (A) Electronic focusing and (B) beam-steering of ultrasonic rays using multiple elements and spatial array processing (images adapted from [80]).

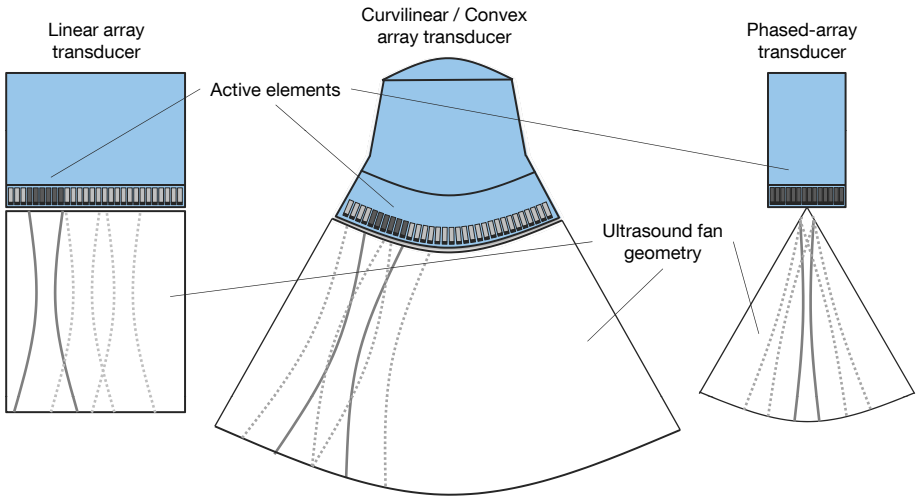

Figure 2.3: Common transducer geometries in medical ultrasound. Linear arrays are used for e.g. vessel imaging while curvilinear arrays are used for e.g. abdominal imaging. Transcranial ultrasound is typically performed using phased-array transducers due to their small footprint (image adapted from [80]).

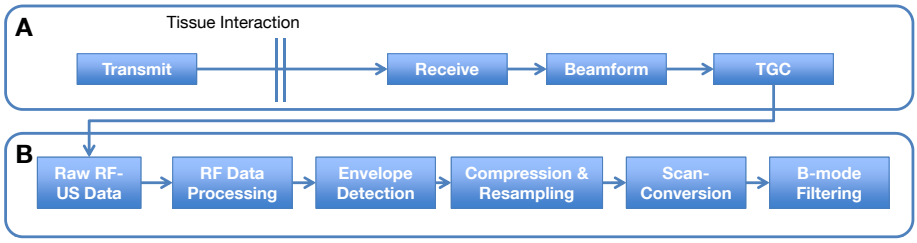

Figure 2.4: Digital signal processing steps necessary to create a clinically usable B-mode image from raw RF ultrasound data.

2.3 RF and B-Mode Ultrasound

As discussed in the last section, ultrasound transducers contain an array of piezo-elements which serve as pistons transmitting acoustic waves into human tissue. Through electronic focusing and beam-steering techniques, the transducer produces narrow beams of high acoustic energy. These beams are used to sequentially insonify narrow strips in front of the transducer. The returning echo are detected as so-called scanlines by the ultrasound system.

By cycling through sub-groups of array elements or through adaptation of delay patterns in transfer- and receive-mode, the beam can be steered in angular direction to create a sweep of scanlines. The received and beam-formed scan-lines are digitized and stored in a matrix with scan-lines as columns and longitudinal echo intensity samples as rows. Rows are increasingly amplified depending on their depth, in order to compensate for the increasing absorption of ultrasound energy in the tissue. This process in called time-gain compensation (TGC). The data matrix after beam-forming and TGC is called radio-frequency (RF) or raw ultrasound data (RF-US). The steps for generation of RF-US data explained until now are visualized in figure 2.4A.

Numerous analogue and digital signal processing steps, which will be explained in the following, are necessary to create a clinically interpretable ultrasound image from the RF-US echo amplitudes acquired at the transducer front face. Figure 2.4B displays the individual signal processing components at each step.

RF data processing: The first digital processing step on received RF scanlines is to perform a bandpass filtering in order to reduce noise and limit the signal to the band of clinically relevant signal frequencies. Bandpass filters can be implemented using a sequence of finite-impulse-response (FIR) filters such as a low-pass and a high-pass filter in sequence, or a direct band-pass implementation, e.g. using Butterworth filters [7]. Further processing steps may be implemented in a clinical ultrasound device, however, the exact filtering steps are often proprietary.

Envelope detection: Each column in a RF data matrix represents a single scanline after beamforming, TGC and bandpass filtering. The received ultra-

sound signal is still modulated with the frequency of the carrier wave and needs to be demodulated and the envelope of the signal has to be found. As described by Misaridis in [106], the amplitude-modulated RF signal can be expressed as a real-valued scalar signal function $f(x)$ in the spatial domain:

$$f(x) = A(x)cos\left[2\pi f_0 x + \phi\left(x\right)\right], \tag{2.10}$$

where $A(x)$ is the amplitude-modulation function, f_0 is the base frequency of the carrier wave and $\phi\left(x\right)$ is the phase modulation function.

For signal processing, and in particular envelope detection and demodulation, it is advantageous to work with a complex-valued signal representation called *analytic signal*. The analytic signal better represents a signal's structural information by decomposing it into its local amplitude and local phase [164]. For example, it has been demonstrated that by using the analytic signal and its structural decomposition, local phase and intensity information can be used for improved non-rigid registration of ultrasound images [177]. The analytic signal $f_A(x)$ of a signal $f(x)$ is defined as:

$$f_A(x) = f(x) - if_H(x) \tag{2.11}$$

where $f_H(x)$ is the Hilbert transform of the real-valued signal $f(x)$. The Hilbert transform is defined in spatial and frequency domain as [177]:

$$f_H(x) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{f(\tau)d\tau}{\tau - x} \Longleftrightarrow F_H(\omega) = F(\omega)\cdot j\cdot \text{sign}(\omega) \tag{2.12}$$

From the definition of the Hilbert transform in frequency domain, one can see that in frequency space, the negative-valued frequency components of the analytic signal cancel each other out, and in the range of positive frequencies, the spectrum is twice the spectrum of the original signal. In frequency domain, we can thus denote the analytic signal as [177]:

$$F_A(\omega) = F(\omega)\cdot[1 + \text{sign}(\omega)] \tag{2.13}$$

Using this definition, the local phase of the signal $f(x)$ can be computed as:

$$\Phi(x) = \tan^{-1}\left(\frac{f(x)}{f_H(x)}\right) \tag{2.14}$$

The envelope function can be computed as the local energy $A(x)$ of the signal $f(x)$:

$$A_{\text{env}}(x) = A(x) = \|f_A(x)\| = \sqrt{f^2(x) + f_H^2(x)}. \tag{2.15}$$

In practical terms, the envelope function inverts negative-valued samples of the original RF signal and forms a tight envelope around the peaks of the resulting signal.

Compression and re-sampling: The obtained envelope signal still contains high-frequency information of spatial samples in longitudinal beam direction. However, the dynamic range of the signal is too high for visual perception by

humans. For example, in transcranial imaging, a dynamic range of around 45-50dB is recommended ([150], see chapter 1.3). This range of amplitudes has to be compressed into a visually perceptible range. Although medical ultrasound devices use proprietary means for signal intensity compression, one basic step in most devices is a logarithmic compression of the signal, which enhances the differences between low-amplitude image regions. Another necessary compression step is a re-sampling of signal bit-depth, also called quantization. Ultrasound machines typically process signal information in a 12-16bit representation, while displaying on a typical computer monitor requires an 8-bit representation. For this, the enveloped, log-compressed RF data is re-sampled and quantized at an 8-bit resolution for further processing and displaying in the machine. Overall, the compression and re-sampling of image sample intensities I can be summarized as [67]:

$$I_{\text{compressed}} = 255 \left(\frac{\log(I / I_{\text{min}})}{\log(I_{\text{max}} / I_{\text{min}})} \right) \tag{2.16}$$

where I_{min} and I_{max} denote a lower and upper threshold for original signal intensities that should still be considered in the compressed range. The obtained values can be stored as an 8-bit character array for further processing.

Scan conversion: In contrast to linear arrays, where all scanlines face away from the transducer perpendicularly and parallel to each other, curvilinear arrays and phased-array transducers perform an angular variation of the beam direction. After RF scanline have been acquired and processed, they are stored in a matrix array in polar coordinates, i.e. in terms of beam angle (matrix columns) and sample depth (matrix rows). In order to visualize the echo reflections in a spatially correct manner, a transformation from polar coordinates to Euclidean coordinates has to be performed. The necessary information can be calculated if exact transducer specifications and scan settings are available. After polar transformation, image pixels falling between rays have to be interpolated, e.g. by using bi-cubic interpolation [67].

B-mode post-processing: After all steps described above, a so-called ultrasound B-mode image is obtained. Typically, however, clinical ultrasound machines still perform several post-processing steps on the grey-valued images in order to further improve visual perception of the image for the physician. The overall goal is to suppress artifacts without diagnostic value while highlighting image features and artifacts which carry diagnostic information. Example operations performed after scan-conversion can include contrast-enhancement [67], edge-enhancement as well as global or local smoothing methods, such as speckle-constrained denoising [91, 153] and other despeckling techniques. For surveys on speckle reduction methods, we refer the interested reader to [148, 86].

2.4 3D Ultrasound

The past few chapters explained the physical background of 2D ultrasound imaging and the B-mode image formation process. In this thesis, we make use of an extension from 2D to 3D ultrasound to investigate potential improvements of the 2D method. 3D ultrasound is a relatively young medical imaging technique, but due to its many advantages has already found applications in several clinical areas such as abdomen, pelvis and breast imaging, cardiovascular applications and ultrasound-guided interventions [110].

There are several technologies [123] for acquisition of 3D ultrasound volumes, based on two basic approaches. The first approach is to compound a 3D volume by acquiring a set of 2D images along with information about the 3D spatial position and orientation of each ultrasound image in 3D space. The 2D images and 3D pose are combined to reconstruct a volume, e.g. through forward or backward *compounding techniques* [175]. The second fundamental approach is to use a *2-dimensional array* of transducer elements and perform electronic beam-steering in all three volumetric directions, instead of only in a plane.

3D ultrasound through spatial compounding: For the first approach, it is necessary to obtain the 3D pose of the transducer for every obtained 2D image. Several solutions exist for this problem. One way is to mechanically steer the transducer element array, e.g. using a micro-step-motor. This approach is used in so-called "*wobbler transducers*" by steering the array in a tilt motion (see figure 2.5A), and creating e.g. a cone-shaped volume. The advantages of this approach are that the system is calibrated by design, the micro-step-motor delivers highly accurate 3D pose information and the mechanic parts are encased together with the imaging array in the transducer's housing, making on-site calibrations unnecessary. However, the footprint of wobbler transducers is usually large, effectively limiting the maneuverability of the transducer. This limits wobbler transducer to special anatomies, such as fetal imaging or abdominal imaging e.g. of the liver [123]. Currently, we are not aware of clinically available transcranial wobbler transducers, probably due to the conflict between footprint of wobbler transducers compared to the small size of the pre-auricular bone window for TCUS imaging, which limits the usability of this approach for our work.

Another approach for 3D pose retrieval is to use an external tracking device, e.g. an electromagnetic (EM) or optical tracker. As long as the tracking device maintains good tracking conditions (e.g. keeping a small distance from the field generator in EM tracking or maintaining an uninterrupted line-of-sight between camera and tracking target in optical tracking), the examining physician is able to change the transducer pose in an arbitrary direction (see figure 2.5B), which is why this approach is coined "*3D freehand ultrasound*". The 3D Freehand method combines the advantages of superior image resolution of a regular transcranial transducer and the ability to perform arbitrary transducer motions during acquisition, which is why we utilize 3D Freehand sonography in this thesis (see chapters 3.1 for a description of our recording setup and 3.2 for the

necessary calibration procedures). The 3D freehand ultrasound technique has also been used in other fields of neurological and neurosurgical research, e.g. for intra-operative guidance [156, 128], ultrasound-based brain shift analysis [129] or transcranial color-Doppler based detection of vascular abnormalities, i.e. cerebral aneurysms [90].

Transducers with 2D matrix arrays: The other, fundamentally different method for 3D scanning is to utilize a 2D matrix array of transducer elements. Similar to a 1D-array, which is useful for electronic steering of focused ultrasound beams in a 2D plane, a 2D array can be used for electronic beam-steering in 3D and creating inherently three-dimensional B-mode scans [67]. Several thousand transducer elements are necessary for proper 3D beam-steering, making this technique technologically much more complex and ultrasound devices significantly more expensive. Also, up to now, the volumetric resolution of 2D matrix array scans is lower than for regular 1D transducer arrays, making its usage inside a 3D freehand ultrasound system a more practical choice for 3D-TCUS scanning. Nevertheless, the 2D matrix array technology bears many advantages, such as inherent 3D scanning without the necessity for a reconstruction step. One of the most significant advantages is that 3D volumes can be obtained much faster, allowing for high-frequency volumetric scans at up to 60 frames per second. This makes in-vivo examination of the beating heart possible, enabling detailed analysis and modeling of the heart, in particular when combined with other modalities such as 4D CT [74].

In summary, 3D Freehand offers the best resolution and is more suited for static TCUS tissue imaging bi-laterally than e.g. wobbler arrays or 2D matrix arrays. However, it is important to note that the methods proposed in the following chapters can also be applied onto other means of 3D B-mode US acquisition such as 2D matrix arrays. Therefore, the methods are complementary to current state-of-the-art as well as to future developments of these techniques.

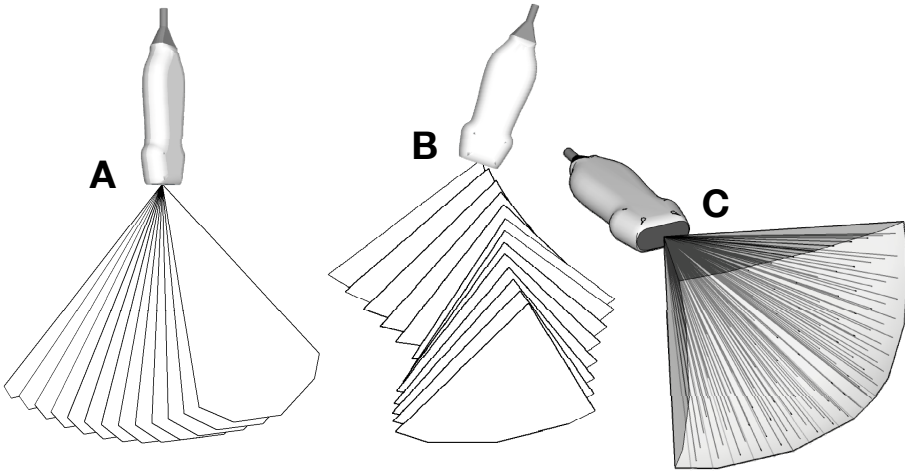

Figure 2.5: Different approaches and acquisition geometries for 3D scanning: A) mechanically steered "wobbler" array, B) 3D freehand ultrasound scan and C) 2D matrix arrays with electronic beamsteering.

2.5 Ultrasound Image Simulation

Based on the physical properties explained in the last sections, ultrasound wave propagation, tissue interaction and ultrasound images as a whole can be simulated, e.g. for purposes of ultrasound training [28], multi-modal registration [174, 136, 94] or ultrasound system design [78].

Towards the end of this thesis, in chapter 7, we will utilize MRI volumes, cropped around the midbrain region, to simulate 3DUS volumes and create a gold-standard dataset for experiments towards improved 3DUS reconstruction. The requirements for the simulation are low in terms of image quality, but the speed needs to be high, due to the necessity to simulate in 3D.

In the following, we describe three main types of ultrasound simulation, namely ray-based, wave-based or convolution-based simulation. Based on this, we will argue why we use convolution-based simulation in our experiments.

Ray-based simulation uses the analogy of optical rays and takes into account their respective physics [84]. This simulation method has been used for multi-modal registration of US to pre-operative CT [174, 136, 94] as well as for fast simulation during ultrasound training [28].

The basic principle is to approximate a desired "tissue" area by a digital map of acoustic impedance regions representing different tissue types. From equation 2.4, we recall that impedance is a function of tissue density and speed of sound, i.e. $Z = \rho \cdot c$. Since ultrasound machines assume a constant speed of sound in tissue, impedance becomes directly dependent on the tissue density. Such information can be gained from other imaging modalities such as CT, where the assumption is made that tissue density correlates with the Hounsfield unit quantifying the amount of X-ray radiation absorption [174]. Therefore, the image intensities in a given CT volume can serve as a relative impedance map for ray-based ultrasound simulation. These intensities are re-sampled along rays representing the ultrasound scanlines. Recalling equations 2.5 and 2.6 for acoustic reflection and transmission percentages, we can denote the reflected sound intensities I_r^k and transmitted sound intensities I_t^k for the k-th sample along the ray as [94]:

$$I_r^k = I_i^k \left(\frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2 \tag{2.17}$$

$$I_t^k = I_i^k \frac{4 Z_1 Z_2}{(Z_1 + Z_2)^2}, \tag{2.18}$$

where I_i^k is the incoming sound intensity, Z_1 the previous and Z_2 the next impedance at the k-th sample, in direction from transducer apex to scan-line end. Given these prerequisites, reflection and transmission maps can be calculated for the tissue and subsequently scan-converted and blended into an ultrasound-approximating image using a recursive calculation of image intensities at each sample k [94]:

$$I_i^k = \begin{cases} I_t^{k-1} & : k > 0 \\ 1 & : k = 0, \end{cases} \tag{2.19}$$

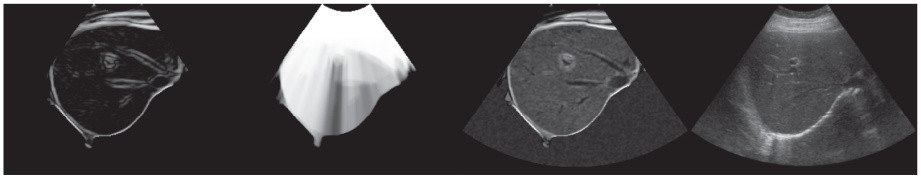

Figure 2.6: Example image for ray-based ultrasound simulation of the liver, simulated from a CT angiography volume, showing reflection map (left), transmission map (middle left), the fused image and ray-based simulation (middle right) and an original ultrasound image of the same anatomy (right) (image reproduced from PhD thesis of Wolfgang Wein [173], by courtesy of the author, email for approval received 24. August 2012).

An example simulation image [173] can be seen in figure 2.6. A highly accelerated implementation for solving this recursive calculation utilizing modern-PC Graphics Processing Units (GPU) has been described by Kutter et al. in [94]. It should be noted that ray-based approaches usually aim at modeling only the major ultrasound-tissue interaction principles, such as reflection and transmission. Refraction usually is ignored since this would require a refraction coefficient map with the same size of the volume or image. Also, refraction only has a minor effect on ultrasound image appearance [94]. Furthermore, ray-based simulations are not capable of simulating fine-scale sound-tissue interactions such as scattering. For incorporation of scattering phenomena and simulating speckle patterns, other approaches have to be used, such as semi-transparent overlays of speckle maps precomputed through wave-based simulation [136] or through learned texture maps from a database of speckle-rich US images [182].

Wave-based simulation allows for simulation of more complex ultrasound phenomena such as interference, scattering and diffraction [84]. The basic principle of wave-based simulation is to simulate wave-front propagations through tissue, emanating from a virtual ultrasound transducer that is simulated in detail as well, including individual piezo-elements, their apertures and beam-steering mechanisms (see chapter 2.2).

Once an ultrasound pulse has been emitted using the simulated transducer model, the wave propagates through tissue and is reflected and scattered by inhomogenities of density and speed of sound. Similar to the ray-based approach, the received field can be approximated or simulated if quantitative values for the underlying scatterer field are assumed. In contrast to ray-based simulation, wave-based simulation is based on solving an appropriate wave equation [78].

A linear model for acoustic wave propagation in an inviscid fluid is given by:

$$\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0 \tag{2.20}$$

where p is the acoustic pressure of the wave and c_0 the equilibrium speed of

sound.

In order to account for changes of wave propagation due to variations in the medium concerning e.g. density and diffusivity, as well as non-linear effects, more sophisticated, non-linear models of wave propagation are often considered. In [116], Pinton et al. describe a non-linear full-wave equation for wave propagation in an attenuating medium as:

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{\delta}{c_0^4}\frac{\partial^3 p}{\partial t^3} + \frac{\beta}{\rho c_0^4}\frac{\partial^2 p}{\partial t^2} + \frac{1}{\rho}\nabla^2 p - \sum_{m=1}^{\nu} \xi_m = 0 \tag{2.21}$$

The first two terms are equal to the terms of the linear wave equation, again with p standing for acoustic pressure and c_0 for equilibrium speed of sound. The third term models thermoviscous diffusivity, with a diffusivity parameter δ, which can be expressed as a function of the absorption coefficient α as $\delta = 2\alpha c^3/\omega^2$, with ω being the angular frequency. The fourth term models nonlinearity, with a non-linearity parameter β and ρ standing for the medium's density. The fifth term models variations in density of the insonified medium. The last ν terms model relaxation mechanisms following the condition:

$$\dot{\xi_m} + \omega_m \xi_m = a_m \omega_m \frac{\Delta c}{c_0}\nabla^2 p. \tag{2.22}$$

This relaxation equation depends on a selected frequency-dependent attenuation law, which models ν peaks at characteristic frequencies ω_m with weight a_m. The changes in speed of sound Δc must adhere to the Kramers-Konig relation to preserve causality [116].

For simulation, it is important to note that the material parameters c_0, δ, ρ and β can be functions of space [116]. In practice, this means that similar to ray-based simulation, a spatial grid of material parameters can be modeled, representing a region of virtual human tissue. Sound propagation through this tissue region can be simulated by using equation 2.21 and applying a numerical solving mechanism such as finite differences [84] to approximate a solution. Different solving mechanisms are available in literature, e.g. using a fast GPU-based solving of the first four terms in equation 2.21 [84]. A linear wave propagation model was used to create "FIELD II", a complex software package which is now widely used [116] for complete simulation of arbitrary ultrasound transducer geometries and excitation, allowing for simulation-guided design of whole ultrasound systems [79].

Wave-based simulations offer high-fidelity ultrasound image results with a high resemblance to real-life ultrasound images and speckle, since effects such as reverberation, dynamic focusing, apodization and micro-scattering are all accounted for by the wave propagation model. However, the solving mechanism is computationally very expensive, in particular for 3D simulations, with high requirements concerning processor power and available memory.

For example, Pinton et al. [116] achieve a highly realistic simulation of wave propagation in 3D on a digital tissue phantom based on histological scans from a section of the abdominal wall, wiht a ground area of approximately 6x4 cm. Insonification was simulated at 2.1MHz, assuming point scatterers of 40 μm diameter and a 5% speed of sound deviation in virtual tissue from the

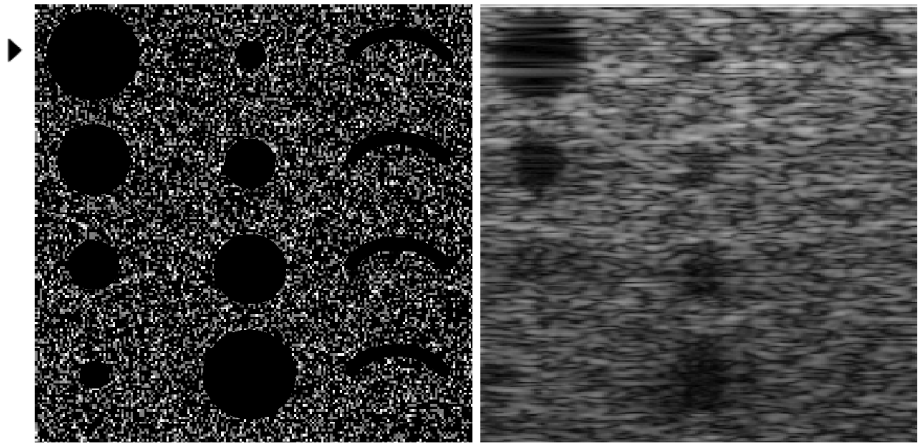

Figure 2.7: Fast, wave-based ultrasound simulation of a digital scattering phantom, based on a GPU-accelerated solver for the Westervelt wave propagation model (image reproduced by courtesy of Karamalis et al. [84]).

assumed average speed of 1540 m/s. Calculation on 56-processor Linux cluster with 64bit architecture required 90GB RAM and approximately 32 hours of run-time [116]. Single 2D-image simulations using Field II take on the order of 10 hours and more [79], while the GPU-accelerated simulation of Karamalis et al. [84] takes approximately 80 minutes for a 2D image simulation. In summary, wave-based ultrasound image simulation creates excellent approximations of real-life ultrasound images (see figure 2.7), but at the cost of high computational demands, making this technique not very practical for 3D simulations.

Convolution-based simulation: A third basic approach for ultrasound simulation is based on a convolution of an artificial scatterer map with a sinc-shaped point spread function (PSF), which simulates the system response of the US system in a simplified manner. The basic principle was proposed by Bamber et al. in [8]. Since then, different convolution-based methods have been proposed, e.g. based on fast convolution in ultrasound k-space (FUSK) [69], or fast 2D/3D simulation based on sequential 1D convolutions (COLE) with a space-invariant PSD for echocardiography [59]. The resulting simulations lack reverberation phenomena and artifacts such as shadowing, but still create visually realistic ultrasound images, in particular concerning speckle appearance, as can be seen in figure 2.8. Most importantly, however, the speed-up of computation compared to wave-based simulation is dramatic. A comparison between FUSK, COLE and Field II for 3D simulation on a small virtual tissue sample conducted by Gao et al. [60], for example, yielded a speed-up on the order of 1000, reducing computation time from approximately 12 hours to 25 seconds using COLE.

As mentioned, in chapter 7.3.3, we will describe how we use a convolution-based simulation method to create crude but fast simulations of 3D ultrasound of the midbrain region from edge-filtered MRI, which allow us to investigate

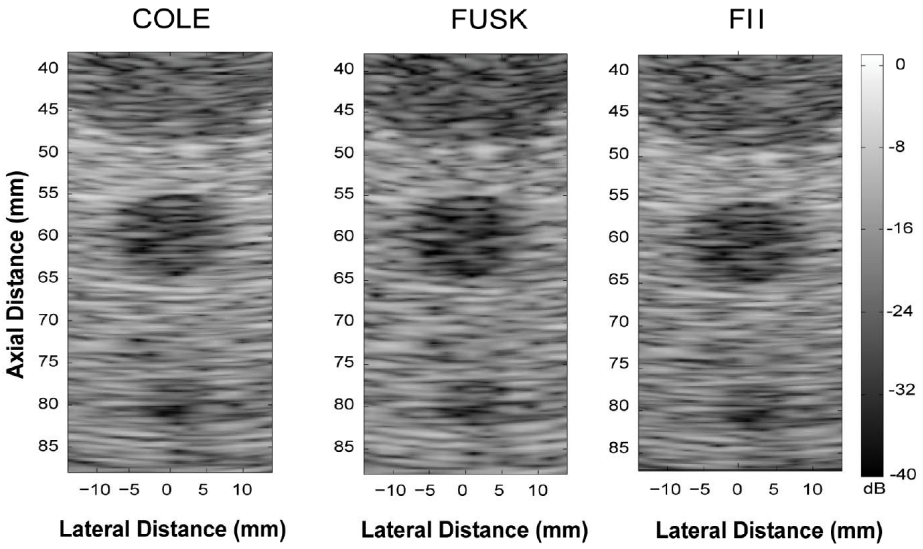

Figure 2.8: Fast, convolution-based ultrasound simulations of a digital scattering phantom, based on convolution of the scattering particles with a convolution kernel approximating the emitted wave pulse. The image shows a comparison between convolution-based methods COLE [59] (left), FUSK [69] (middle) and wave-based simulation using Field II [78] (right). (image reproduced from Gao et al. [60]).

a novel and advanced reconstruction method for bi-lateral ultrasound of the midbrain in chapter 7. We utilize convolution-based simulation for two reasons: 1) The quality and realism of simulated images are sufficiently high for a region-based segmentation algorithm to pick up anatomic boundaries and 2) the speed of simulation is faster by three magnitudes.

3 Data Acquisition

3.1 Bi-lateral Transcranial 3D Freehand Ultrasound Setup

As described in section 2.4, the currently best way for obtaining 3D transcranial ultrasound (3D-TCUS) image volumes is to use the "3D Freehand Ultrasound" approach. In that section, we also described the general setup for volumetric ultrasound acquisition in a 3D Freehand system. In the data acquisitions for our main study in this thesis, we used a clinical ultrasound scanner (Siemens Acuson Antares, Siemens AG, Erlangen, Germany) with a 2-4 MHz probe. Attached to the probe was a tracking target for detection by a clinically approved optical tracking device (NDI Polaris Spectra; Northern Digital Inc, Waterloo, ON, Canada). The 2D B-Mode or color Doppler images were duplicated from the ultrasound scanner using a high-resolution frame-grabber card (Unigraf UFG-05; Unigraf Inc., Espoo, Finland). An ultrasound probe calibration step using the Single Wall method was performed prior to the image acquisition, a more detailed explanation of the calibration procedure is given in section 3.2.2. Additional to the regular 3D Freehand setup, which includes a tracking target for the ultrasound probe, we recorded the position of the subject's head with an additional optical tracking target attached to the forehead, as illustrated in Fig. 3.1. First, this allowed for compensating any head motion during acquisition and second, it allowed us to reconstruct ultrasound scans from different positions with respect to the head into one single volume. This enabled a reconstruction of 3D ultrasound sweeps from the left and right bone windows into one bi-lateral volume.

Throughout the rest of this thesis, we will use the following notation for transformation matrices:

$$C_{target} = {}^{target}T_{source} \cdot C_{source}, \tag{3.1}$$

where C_{target} denotes the target coordinate system, C_{source} the source coordinate system and ${}^{target}T_{source}$ denotes a transformation *from* the source *into* the target

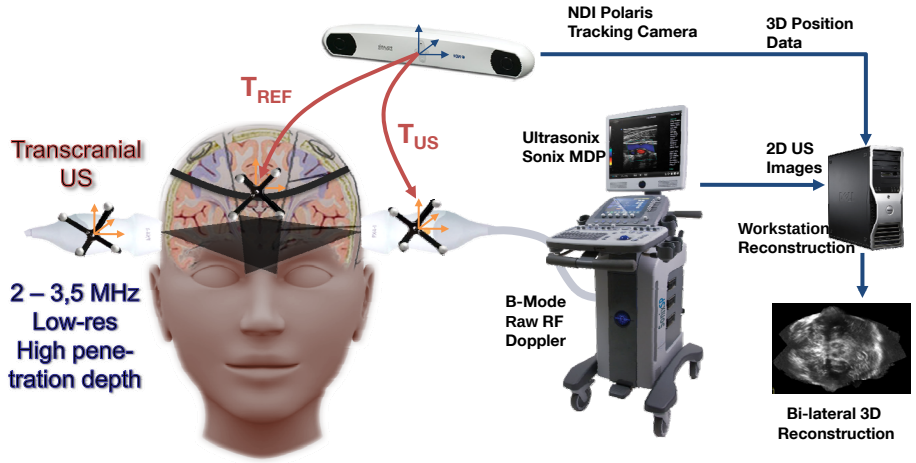

Figure 3.1: Setup for the bi-lateral 3D freehand ultrasound acquisition transcranially.

coordinate system. Concatenations of matrices mathematically describe linear transformation chains along different coordinate system.

As illustrated in Fig. 3.1, the main transformations of interest in the bi-lateral freehand US system are the $4x4$ tracking matrices ${}^{US}T_{World}$ of the ultrasound transducer and ${}^{REF}T_{World}$ of the reference target attached to the subject's forehead. Please note that the tracking targets US and REF move rigidly with respect to each other if the patient head moves during acquisition. Hence, head motion is compensated if the ultrasound probe tracking data is used in the coordinate system of the reference target.

A detailed breakdown of the transformation steps for transducer calibration and 3D volume reconstruction are given in chapters 3.2.2 and 3.3, respectively.

3.2 Ultrasound Probe Calibration

3.2.1 3D Freehand Calibration Procedures in Literature

In order to obtain the spatial relation between the tracking target mounted to the ultrasound probe and the 2D ultrasound image floating in 3D space in front of the transducer, an ultrasound probe calibration step has to be performed, before 3D acquisitions can be performed.

In general, the probe calibration step consists of sequential scans of a gel- or water-immersed calibration phantom with known geometric properties, during which the ultrasound probe is continuously tracked. The appearance of the phantom geometry in the 2D ultrasound images can then be detected in order to obtain 2D pixel locations of relevant geometric phantom features. Depending on the phantom design, a mathematical dependency containing the desired calibration matrix can be formulated between the 3D pose of the tracking target and the appearance of phantom features in the 2D ultrasound image. Once a sufficient set of 3D transducer poses and corresponding 2D pixel coordinates for features in the US image has been collected, the desired calibration matrix can be solved for, e.g. using least-squares approximation or other optimization techniques.

In a Master thesis supervised during this PhD thesis (see appendix section B.4), the literature on ultrasound calibration has been extensively reviewed and summarized [38]. Further surveys of 3D freehand ultrasound calibration methods can be found in [104] and [71]. A wide range of calibration phantom geometries exist, along with suitable feature detection methods as well as mathematical correspondences to solve for the calibration matrix. The most widely used phantom geometries are displayed in figure 3.2. Depending on the geometry of the phantom, features in the 2D US image can range from single intersection points (single cross-line or three-wires phantom), point sets (z-phantom) or pre-defined feature points such as corners or edges (2D shape alignment phantom) to complete intersection lines (single-plane, membrane or Cambridge phantom).

Each phantom design has its own advantages and drawbacks in terms of construction complexity, usability, complexity of automatic segmentation, quickness and reported accuracy. A comparative overview over these parameters was assembled in [38], the results are given in table 3.1. In this thesis, we utilized a single-wall phantom with a nylon membrane for probe calibration, since it is comparatively cheap and easy to construct, while offering an acceptable accuracy which is reported to be in the mid-range of all calibration methods [38]. Since we calibrate the probe previous to acquisition and since our target is rigidly fixed to the ultrasound transducer, calibration has to be performed only once and thus the comparatively long duration required for calibration when using a single-wall phantom (approximately 15-30 minutes) can be neglected.

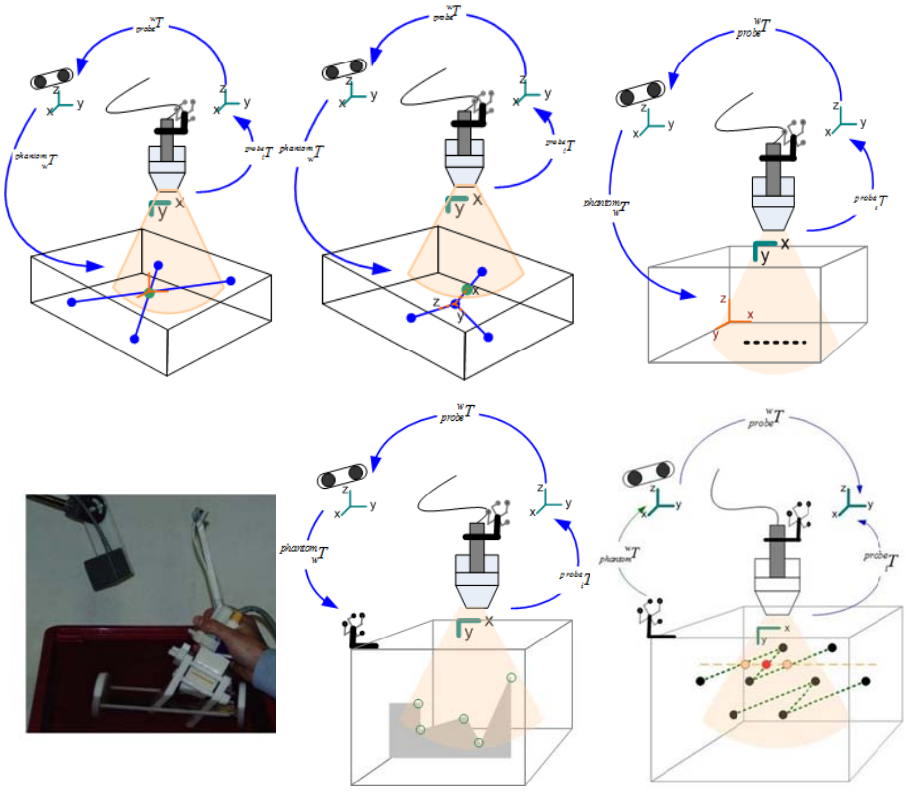

Figure 3.2: Illustration for various calibration methods, summarized in the Master thesis of Mei Chuan Chen (supervised in this PhD thesis).

	Easy to construct	Easy of usability	Automatic segmentation	Calibration speed	Accuracy (mm)
Single cross-wire	+	–	–	-	1.12 + 0.04
Three-wire	-	+	–	-	2.18 + 0.15
Single wall	++	+	+	-	1.63 + 0.04
Cambridge	-	+	+	+	1.33 + 0.23
2D shape alignment	-	-	-	-	+ 2.28
z-phantom	+	++	+-	+	1.5 + 2.46

Table 3.1: Advantages, drawbacks and accuracies of US probe calibration methods in related literature (from [38]).

3.2.2 Single-wall Membrane Phantom Calibration

For 3D freehand ultrasound probe calibration using a single-wall phantom, a water-immersed nylon plane is scanned from various probe positions and angles. The transformation relevant for calculation of the desired calibration matrix are depicted in figure 3.3. As mentioned before, in terms of notation, please note that a transformation matrix denoted as ${}^{B}T_{A}$ represents a rigid $4x4$ transformation matrix from coordinate system A into coordinate system B. Each rigid transformation contains six parameters, three for Euler rotation angles (ϕ, θ, ψ) and three for translation (x, y, z). From these, the transformation ${}^{B}T_{A}$ can be calculated as:

$$
\begin{aligned}
&{}^{B}T_{A}(\phi,\theta,\psi,x,y,z) = \\
&\begin{bmatrix}
\cos\theta\cos\psi & -\cos\phi\sin\psi+\sin\phi\sin\theta\cos\psi & \sin\phi\sin\psi+\cos\phi\sin\theta\cos\psi & x \\
\cos\theta\sin\psi & \cos\phi\cos\psi+\sin\phi\sin\theta\sin\psi & -\sin\phi\cos\psi+\cos\phi\sin\theta\sin\psi & y \\
-\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta & z \\
0 & 0 & 0 & 1
\end{bmatrix}
\end{aligned}
\tag{3.2}
$$

When scanning the phantom, the intersection of the ultrasound image plane with the phantom plane results in a line intersection. Two points p_1 and p_2 can be defined for each intersection line l_i, where each point has image coordinates $(p_{i,x}, p_{i,y})$, and x, y represent the column and row indices of the intersection point in pixel coordinates. The y-coordinate (or row index) of the points has to be corrected for speed of sound of ultrasound in water. According to Marczak [101], the speed of ultrasound in a water bath of temperature T $[K]$ can be calculated as:

$$
\begin{aligned}
c(T) = ((((2.787860 \cdot 10^{-9} \cdot T - 1.398845 \cdot 10^{-6}) \cdot T + \\
3.287156 \cdot 10^{-4}) \cdot T - \\
5.799136 \cdot 10^{-2}) \cdot T + \\
5.038813) \cdot T + \\
1.402385 \cdot 10^{3};
\end{aligned}
\tag{3.3}
$$

Please note that once calibration has been performed using the single-wall phantom and including this speed-of-sound compensation, it is not important which speed of the sound any future medium will have. Once the geometric configuration of the transducer tracking target and the origin of the image plane have been determined, this geometry will stay rigid, independent of the scanned medium in front of the transducer.

The complete transformation chain, transforming a 2D line point from the ultrasound image onto its physical location $(p^{F}_{i,x}, p^{F}_{i,y})$ on the calibration plane can be denoted as:

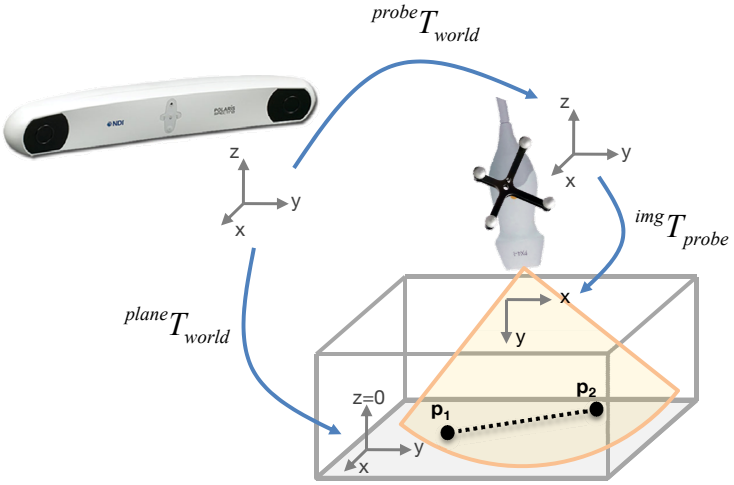

Figure 3.3: Transformation matrices for single-wall calibration of a 3D freehand ultrasound transducer.

$$ {}^{PhantomPlane}T_{Tracker} \cdot {}^{Transducer}T^{-1}_{Tracker} \cdot {}^{USPlane}T^{-1}_{Transducer} \cdot T_{Scale} \begin{pmatrix} p_{i,x} \\ p_{i,y} \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} p^F_{i,x} \\ p^F_{i,y} \\ p^F_{i,z} = 0 \\ 1 \end{pmatrix} \tag{3.4} $$

The constraint in this equation is that line intersection points have a zero z-value in the phantom coordinate system. The two (arbitrarily) selected line intersection points for each image yield two projections. Since equation 3.4 cannot be solved analytically, the following cost function can be defined [71]:

$$ f_{plane} = \sum_{N}^{i=1} \left((p^F_{1i_z})^2 + (p^F_{2i_z})^2 \right) \tag{3.5} $$

In the equation above, N denotes the number of acquired phantom images. If the pixel scalings in $x-$ and $y-$direction are previously determined (e.g. by several line measurements using the US machine's built-in measuring tool, or through a SDK research interface to the machine), equation 3.5 becomes a function of nine variables, six for the rigid calibration matrix and three for the unknown calibration plane (since we assume the plane to have infinite extent, we do not care for its x/y-offset or for the rotation around its plane normal, but we only need to solve for two rotational offsets and one z-offset from the world coordinate system). By minimizing the cost defined in equation 3.5, we thus simultaneously solve for both the desired calibration as well as for the location of the phantom plane in world coordinates. The minimization can be performed e.g. with a Levenberg-Marquardt optimizer.

In terms of implementation, we used the "UltrasoundCalibration" application plugin provided by the CAMPAR framework of the CAMP group [146]

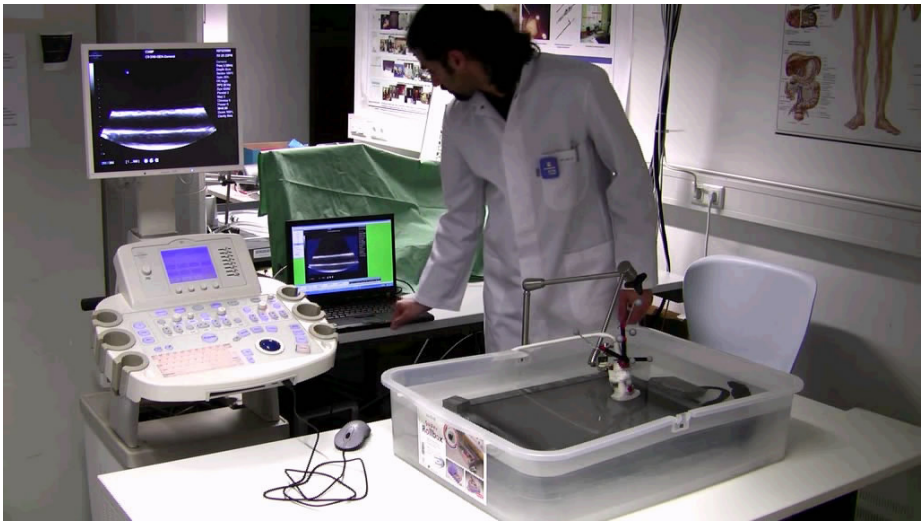

Figure 3.4: Physical setup of the ultrasound probe calibration procedure.

and later implemented an independent US calibration program using MATLAB. The CAMPAR plugin features synchronized data stream acquisition from the US machine and the 3D tracker, the logical and mathematical backend and solver as well as a UI front-end to perform all necessary steps for ultrasound probe calibration. The steps include a tool for ultrasound fan geometry setting, automatic line detection, speed-of-sound compensation given a measured water bath temperature, a pixel scaling tool, temporal and spatial calibration as well as a simple evaluation tool. Evaluation is performed qualitatively and once a calibration matrix has been calculated. The calibration quality is assessed by re-projecting the virtual plane into the current 2D US image, given the current 3D pose of the transducer. The intersection is calculated in pixel space and overlayed onto the US image, where it can be compared to the actual visible intersection line. An image of the physical calibration setup with a 6 degree-of-freedom (DOF) mechanical arm holding the ultrasound probe, the ultrasound machine, tracking system and calibration computer can be seen in figure 3.4.

3.3 3D Ultrasound Reconstruction

Temporal synchronization of image and tracking data: During 3D Freehand acquisition, two streams of data are collected, one containing ultrasound B-mode or RF images and the other one containing 3D image pose. Since these data streams come from different, unsynchronized devices, the data pairs have to be synchronized using acquisition timestamps. In order to obtain one tracking datum per US image, each US image is associated with its two nearest tracking poses using the recorded timestamps. The exact tracking pose for each US image is then obtained using a linear interpolation between the two. Linear interpolation for translation is straightforward. For the rotational part, quaternions are interpolated using spherical linear interpolation (SLERP) [144].

Selection of ROI and bounding box: Once synchronized pairs of images and 3D poses have been obtained, the 3D locations of all 2D pixel in world coordinates can be calculated using the transformation chain in equation 3.6. Before reconstruction, the region-of-interest (ROI) for reconstruction can be determined by calculating the corner-points of all 2D images floating in 3D space and positioning a bounding box around all images. The bounding box orientation can also be pre-determined e.g. by selecting one of the 2D images as a reference slice. This is helpful to orient the 3D volume along a certain desired cut-plane through the anatomy.

Backward compounding: For reconstruction, the bounding box is rasterized in a grid of isotropic voxels. We denote the set of all voxel locations in the grid as $\mathbf{X}$, the 3D position of the i-th voxel as x_i, and the intensity at this voxel location as $I(x_i)$.

Reconstruction is performed using a so-called backward-compounding technique [175], in which the intensity of each 3D voxel is calculated using a mapping of nearby 2D pixel intensities. The concrete implementation used in our study is explained in the PhD thesis of our project partner Tassilo Klein [89], the most important details are explained in the following.

The pixels $p_i = \left(p_{i,x}, p_{i,y}, 0, 1\right)^T$ from each 2D US image can be transformed onto its location in 3D space $p_i^{3D} = \left(p_{i,x}^{3D}, p_{i,y}^{3D}, p_{i,z}^{3D}, 1\right)^T$ using the following transformation chain:

$$\begin{pmatrix} p_{i,x}^{3D} \\ p_{i,y}^{3D} \\ p_{i,z}^{3D} \\ 1 \end{pmatrix} = {}^{World}T_{Transducer}^{-1} \cdot {}^{Transducer}T_{USPlane}^{-1} \cdot T_{Scale} \begin{pmatrix} p_{i,x} \\ p_{i,y} \\ 0 \\ 1 \end{pmatrix} \tag{3.6}$$

By transforming all pixel locations into 3D space, we can associate each pixel to its nearest voxel. Thus, for each voxel x_i, we obtain a set $A(x_i)$ of associated pixels with intensities $I(p_j)$ and distances $d(x_i, p_j) = \left\|x_i - p_j\right\|$ to the voxel center. A maximum distance D can be defined in order to constrain the amount of pixels taken into account to calculate the voxel intensity. This results in the following set of relevant pixel intensities per voxel:

$$A(x_i) = \left\{ \left(d_j(x_i, p_j), I(p_j) \right) | d_j < D; d_j = \left\| x_i - p_j \right\| \right\} \tag{3.7}$$

The calculation of the voxel intensity $I(x_i)$ is performed through a weighting or selection function $f\left(A\left(x_i\right)\right)$. Some common options for the weighting or selection function are [175]:

Nearest Neighbor Selection:
The voxel intensity $I(x_i)$ is set to the intensity of the pixel p_j which is closest to the voxel centroid.

$$f\left(A\left(x_i\right)\right) = I\left(p_j\right) | d_j\left(x_i, p_j\right) = \min\left\{d_j\right\} \tag{3.8}$$

The advantage is that original pixel intensities appear in the final reconstruction, which results intensity distributions that stay true to the original intensity distributions of 2D images. However, most of the acquired 2D US information is discarded when this selection is performed.

Inverse Distance Weighting:
The voxel intensity $I(x_i)$ is calculated by a weighted mean of all associated pixel intensities, the weights for each pixel intensity are calculated as the normalized inverse distance to the voxel centroid [143]:

$$f\left(A\left(x_i\right)\right) = \sum_{n}^{j=1} I\left(p_j\right) \frac{d_j^{-\mu}}{\sum_n^{k=1} d_k^{-\mu}} \tag{3.9}$$

The factor $\mu > 1$ is a smoothness parameter that controls the decay of influence for pixel intensities, with increasing distance.

Gaussian Weighting:
For smooth reconstructions, distances can be weighted using an isotropic 3D Gaussian distance weighting kernel:

$$f\left(A\left(x_i\right)\right) = \frac{\sum_{j=1}^{n} I(p_j) \exp\left(-d_j^2/\sigma^2\right)}{\sum_{j=1}^{n} \exp\left(-d_j^2/\sigma^2\right)}. \tag{3.10}$$

The parameter σ represents the width of the kernel and thus again controls the decay of influence for pixel intensities, with increasing distance.

3.4 Clinical Dataset Acquisition with 23 Subjects

Using the aforementioned Freehand 3D-TCUS recording system, we acquired a dataset with volumetric midbrain and SN acquisition. For quantitative analysis of our dataset as well as for development and validation of segmentation algorithms, a medical expert performed a full, manual and volumetric segmentation of the target region in 3D. In order to compare midbrain appearance of healthy people and Parkinson's Disease (PD) patients, we acquired a small cohort of 11 people for each group. We would like to denote that given the volumetric acquisition, the size of the dataset and the manual expert segmentation on the data, this 3D dataset is unprecedented in the community of diagnostic ultrasound for neurology.

3.4.1 Demographic and Anamnetic Details

We performed data acquisition as described on 23 subjects consisting of 11 PD patients (mean age 65.8 y, s.d.: .9; 4 female/7males) and 12 healthy controls (mean age 55.6 y, s.d.-9.0; 6 female/5males). Patients were diagnosed as idiopathic PD according to the UK Parkinson's Disease Society Brain Bank [72]. On average, subjects from the PD group had a disease duration of 6.4 yrs. (s.d. 3.8) and Hoehn & Yahr scores varied from 1.0-3.5 (mean 2.1, +/-0.8). One healthy subject had to be excluded due to an insufficient bone window, prohibiting proper visualization of the midbrain region (4.3%), which is slightly lower than the expected exclusion rate from 2D transcranial ultrasound examinations (see also chapter 4.4). All subjects participated after given consent. Ethics approval for our study was provided by the Ethical committee of the medical faculty of the Ludwig-Maximilians-University of Munich, Germany. Demographic details are summarized in table 4.1 for healthy controls and in table 4.2 for Parkinson patients.

3.4.2 Ultrasound settings and Sonographer Details

All scans were performed by one investigator experienced with 2D ultrasound of the midbrain (AP). On each temporal side, the US probe was adjusted to screen midbrain structures and was moved slowly to scan the whole volume under consideration of cranial, caudal, anterior and posterior alignments. Between two to four of these scans of approximately 30 seconds length were acquired. This corresponded to a total of 1000 2D image frames for each subject and for each bone window side. The settings of the ultrasound machine were as follows: Tissue Harmonic Imaging (THI), 14cm penetration depth, S-curve shaped time gain compensation. These settings correspond to previously published recommendations [22, 150] (see chapter 1.3), except for the fact that we used THI in our study, due to a visually clearer emphasis of SN echogenicities. Since our evaluation of 3D data was novel, no established reference cut-off values were known and the same settings were applied in all acquisitions, no unwanted bias was introduced.

3.4.3 Manual Segmentation for Establishing Goldstandard

A manual segmentation was performed on the resulting 3DUS volumes by a medical expert[1] blinded for the diagnosis of the patient, in order to obtain training data for the classification algorithm. For this procedure, we used ITK Snap software as segmentation user interface [181]. Segmentation was performed on 12 reconstructed slices (slice thickness: 0.45 mm) in axial direction leading to segmentation of a slab with 5.4 mm thickness. The first slice to be segmented began 1.35 mm caudal to the slice on which the ground of the third ventricle was visible. Subsequent slices were segmented in a caudal direction. In each slice, three areas of interest were segmented, the outline of the midbrain, and outlines of the hyper-echogenic areas within the right and left SN.

Mathematically, we denote the acquired image volume as a 3-dimensional subspace $\Omega_I \rightarrow \mathbb{R}^3$ and the segmentation as a classification function $\mathbf{C} : \Omega_I \rightarrow \mathbf{c}$, which maps each voxel $\mathbf{x} \in \Omega_I$ onto a label $\mathbf{c}$. This label denotes whether the voxel belongs to the midbrain ($\mathbf{c} = 1$), hyper-echogenicities of the right SN ($\mathbf{c} = 2$), hyper-echogenicities of the left SN ($\mathbf{c} = 3$) or to the background ($\mathbf{c} = 0$).

$$\mathbf{C}(\mathbf{x}) = \mathbf{c} \in \{0, 1, 2, 3\}, \text{ where } \mathbf{x} \in \Omega_I \subset \mathbb{R}^3 \tag{3.11}$$

Slice-wise manual segmentation of these detectable structures was followed by a calculation of the corresponding volumes by software. The average duration for manual segmentation of one volumetric was approximately 20 minutes. An example of these scans, combining US image information from both temporal lobes into one volume can be seen in figure 3.5.

[1] Dr. Annika Plate, Neurologische Klinik und Poliklinik, Klinikum der LMU Grosshadern, Munich, Germany

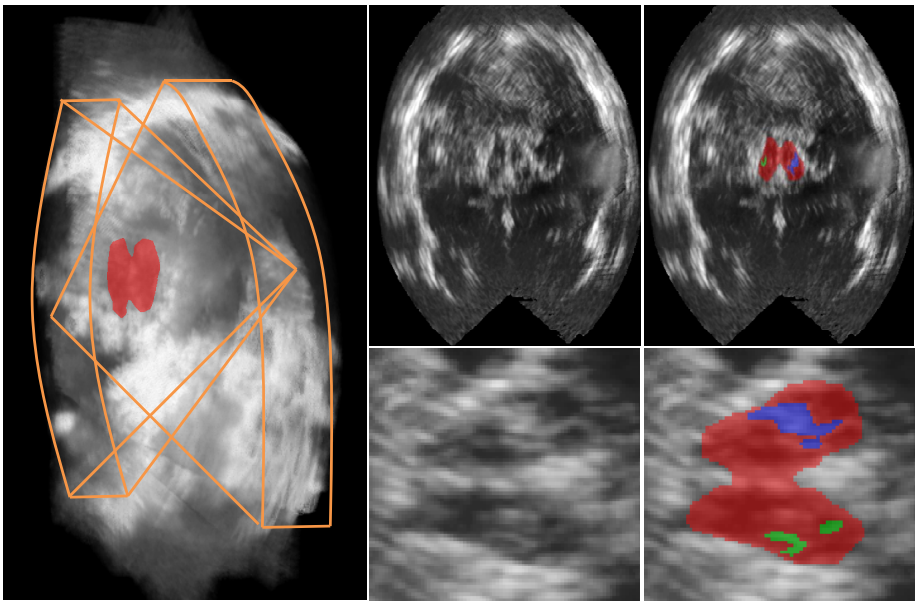

Figure 3.5: Bi-lateral 3D-TCUS reconstruction (left) of one subject from the 22-subject study, a 2D bi-lateral cross-cut through the 3D volume without (top middle) and with manual expert segmentations created in ITK Snap [181] (top right), as well as a magnification of the midbrain area. The midbrain segmentation is highlighted in red, SN right is displayed in green, SN left as blue.

4

Towards Computer Aided Diagnosis of Parkinson's Disease Using 3D TCUS

4.1 Objective and Motivation

In the last chapter, we described our bi-lateral 3D-TCUS system and acquisition setup as well as the acquisition of the first 3D-TCUS dataset for diagnosis of neurological movement disorders together with our clinical partners at the "Neurologische Klinik und Poliklinik" (LMU Klinikum Grosshadern, Munich, Germany). As described, the dataset contains scans from 11 previously diagnosed Parkinson's Disease (PD) patients and 11 healthy controls.

All 3D-TCUS datasets were labeled by a medical expert, who was experienced in the regular 2D technique. In this chapter, we describe the extraction of several single- and multi-dimensional feature sets from these 3D segmentations and automatic classification of subjects into healthy/PD classes.

In the regular 2D-TCUS technique for PD diagnosis, classification is performed by thresholding the area of measured SN hyper-echogenicities (SNEs) in 2D B-mode images [168, 22, 99]. Here, we follow a similar assumption, namely that an excess *volumetric* amount of 3D SNEs is an indicator for PD. Since we measure separate volumes for both sides of SNEs and the midbrain, we also investigate multi-dimensional feature vectors and non-linear classification using machine learning techniques such as Support Vector Machines.

In this chapter[1], the objective is to perform a first quantitative analysis of our 22-subject dataset and investigate for the first time whether 3D measurements can be principally used for computer-aided diagnosis of midbrain abnormalities.

[1]This chapter is based on and extends the paper submission "3D Sonographic Examination of the Midbrain for Computer-Aided Diagnosis of Movement Disorders" to the journal "Ultrasound in Medicine & Biology" [121]

4.2 Materials and Methods

In this chapter, several methods are used towards computer-aided diagnosis. All methods are based on a previous 3D segmentation of and volumetric measurement of left and right SNEs as well as the midbrain volume. Here, we make use of manual expert segmentations (see chapter 3.4.3). Later, in chapters 5 and 6, we will propose methods for automatic segmentation and volume measurement.

For automatic classification of subjects into the classes "healthy control" vs. "PD patient", we experiment with several single- and multi-dimensional features, which are explained in the following. Next, we will explain the single-dimensional classification using ROC analysis and multi-dimensional classification using non-linear Support Vector Machines (SVM). As a final step, we will provide a comparison of both raters using a set of segmentation overlap and correlation measures, as a small but first inter-rater observability analysis on 3D-TCUS data.

4.2.1 Single-dimensional and multi-dimensional feature sets

In order to distinguish PD patients from healthy control subjects, it is necessary to choose distinguishing features from the acquired 3DUS datasets and the manual segmentations. Since no studies on diagnosis of neurological movement disorders using transcranial 3DUS have been reported on so far, we decided to derive sets of single-dimensional and multi-dimensional features from the manual segmentation and compare the classification performance. The following single-dimensional feature sets were evaluated:

- **Larger side 3D SN volume (absolute):** i.e. the volume of the larger of both 3D-segmented SN hyper-echogenicity regions. This measure is given in mm^3.
- **Larger side 3D SN volume (relative):** i.e. the larger of both hyperechogenic SN 3D-regions, divided by the overall volume of the midbrain. The rationale for the midbrain volume division is to normalize the SN echogenicities with the midbrain size, which implicitly represents the brain size and anatomic visibility due to acoustic bone window quality. This measure is given in percent.
- **Sum of 3D SN volumes (absolute):** i.e. the sum of 3D-segmented SN echogenicity volumes in the left and right hemisphere, measured in mm^3.
- **Sum of 3D SN volumes (relative):** i.e. the sum of segmented SN echogenicity volumes in the left and right hemisphere, divided by the overall midbrain volume, similar to above as a brain size and bone window normalization factor.
- **Largest 2D area:** i.e. the largest 2D area of SN echogenicities in any one of the two hemispheres and on a single slice. This feature is comparable to the 2D transcranial sonography method which is currently performed in related work.

The following multi-dimensional feature sets were evaluated:

- **VolumeSNLeft and VolumeSNRight (absolute):** i.e. a 2D-feature vector made up of the volumes of SN echogenicities in the left and right hemispheres.
- **VolumeSNLeft and VolumeSNRight (relative):** i.e. a 2D-feature vector made up of the relative volumes of SN echogenicities in the left and right hemispheres, each normalized as mentioned above by the midbrain volume
- **VolumeSNLeft and VolumeSNRight and VolumeMidbrain (absolute):** i.e. a 3D-feature vector made up of the volumes of SN echogenicities in the left and right hemispheres and the midbrain volume.

The feature sets were used to design classifiers which automatically distinguish between the Parkinson patients (PP) and the healthy controls (HC). Different classification methods were used. For classification of single-dimensional features, we utilized the method of receiver-operating-characteristic (ROC) curve analysis, which allows to find a hard threshold or cut-off for the analysed parameter. For multi-dimensional feature sets, we utilized Support-Vector-Machines (SVM), a non-linear supervised classification method from the machine learning domain. Both methods will be explained in the following sections.

4.2.2 Single-dimensional classification using ROC curves

For single-dimensional feature classification, we perform a receiver-operating-characteristic (ROC) curve analysis [65], which is an often-used tool in medical tests in order to determine an optimal cut-off or threshold value, e.g. for a diagnostically relevant measurement variable [185].

The ROC curve analysis is based on the principle of binary testing. In a binary test, a population of test objects is classified in two classes (e.g. positive vs. negative, healthy vs. diseased). In our case, the test itself is a simple test based on a binary thresholding of one of the single-dimensional features described in the previous section. For example, we classify a subject as PD patient, if the feature *"Larger side 3D SN volume (absolute)"* exceeds a threshold τ. In ROC analysis, the threshold τ is varied along the feature values present in the dataset. Since the class c_i for every subject i is known beforehand, the classification quality can be measured for each τ_i. The classification quality measures can be derived from the confusion matrix, also known as contingency table:

The ROC curve plot visualizes the classification accuracy by assuming the value τ_i as the current threshold and calculating the False Positive Rate (FPR_i) and the True Positive Rate (TPR_i) for this particular test. This is repeated for all feature values τ_i in the dataset. Each pair of classification result values (FPR_i, TPR_i) is plotted in the ROC curve. Furthermore, we can calculate the f-measure (also called F_1-score) for each classification result. The F_1-score is

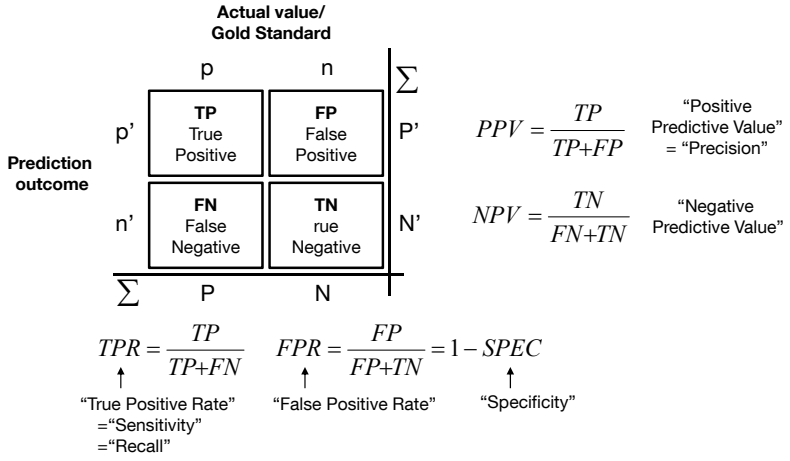

Figure 4.1: Contingency table or confusion matrix, formulating the possible outcomes of a binary test and their derivations.

defined as the harmonic mean between precision and recall of a test. Thus, it serves as a single value describing the test's classification quality:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4.1}$$

As the optimal threshold in the ROC curve analysis, we take the threshold τ_i for which the best F_1-score is obtained. Apart from the optimal threshold τ_i and it's associated values for sensitivity and specificity of the test, we can calculate another measure for the quality or reliability of the ROC curve test, namely the area-under-the-curve value (AUC). It serves as a summary statistic and can be interpreted as the probability that the classifier will rank a randomly selected positive example higher than a randomly selected negative example [54].

4.2.3 Multi-dimensional classification using SVM

We provide the SVM classifier with a sub-set of data to learn the relationship between several volume features of each subject (e.g. volumes of SN left, SN right and of the midbrain) and the subject's class (i.e. healthy vs. Parkinson). In the learning stage, the SVM classifier tries to find a plane which optimally separates the individuals in the multi-dimensional feature space. Non-linear kernel functions and the optimal fitting of a linear plane in higher-dimensional space allow the SVM classifier to find hyper-planes which allow for non-linear separation of the original feature space into classes. In combination with soft margins, SVMs allow for relatively robust and adaptive performance in difficult classification problems. In the test stage, the features of an unknown subject are presented to the SVM, which can then classify the subject into known classes based on the learned feature space separation. We refer the reader to Christianini and Shaw-Taylor [45] for more details on SVM theory, algorithm

parameters and SVM applications.

Several model parameters influence the classification performance of SVMs. The complexity parameter determines the granularity of the hyper-plane, i.e. the hyper-plane smoothness. Another parameter is the kernel choice. In this work, we use Radial Basis Functions (RBF) as the non-linear kernel, with variance Γ. To find the optimal hyper-parameters, a grid search is performed. The SVM complexity and variance parameters C and Γ are varied in logarithmic steps, between $10^-2 : 10.^9$ and $10^-8 : 10^6$ respectively. Since all feature dimensions vary in different ranges, all features are mapped onto the unit d-ball (Hsu and Lin, 2002). The classification experiments are performed in a cross-validation scheme, in order to avoid bias of the classification. We use leave-one-out cross-validation for all experiments. As performance measures for the binary classification, we use sensitivity (SEN), specificity (SPEC) and f-Measure (fM), or F1-score, which is the harmonic mean between precision (PREC) and sensitivity (SEN) (see Fig. 4.1).

4.2.4 Methods for Inter-observer Comparison

In section 3.4.3, we described the manual segmentation by a medical expert. In this section, we want to elaborate on several methods which allow us to give a first estimate for inter-observer variability and reproducibility of the 3D method. It is important to note that in order to be able to perform an inter-observer study, we need at least one more segmentation of our dataset, compared to the segmentation provided by the medical expert, which is why we created a second segmentation of our entire dataset.

As mentioned before, the first observer ($O1$) was a neurologist[2] with two years of clinical practice, including further training in TCUS imaging of the midbrain/SN using the 2D method, as well as experience with 3D-TCUS imaging for PD diagnosis. The second observer ($O2$) was a post-graduate student with non-medical background[3], but prior experience in 3D ultrasound imaging. The second observer was educated and trained for the task in terms of midbrain anatomy as well as typical appearance of midbrain and SN in transcranial ultrasound. Due to differing backgrounds, we denote O1 as an expert observer and O2 as a novice observer for this study. It is important to note that both observers were blinded to the subject's identity and class (healthy versus PD), by encoding subjects with a numeric ID and several months of absence from the dataset between the time of acquisition and the time of segmentation. Furthermore, raters were only exposed to 3D reconstructions, which typically have a different appearance than the 2D images observed during the non-blinded acquisition.

In order to evaluate the inter-rater observability and give a first impression on the objectivity of the 3D method, we calculate four measures, the first two reflect the overlap of manually segmented volume regions (midbrain, SN right, SN left). The latter two measures reflect correlation of volume values measured

[2]Dr. Annika Plate, Neurologische Klinik und Poliklinik, Klinikum der LMU Grosshadern, Munich, Germany

[3]Seyed-Ahmad Ahmadi, thesis author

by observers $O1$ and $O2$.

Volume segmentation overlaps:
Volume segmentation overlaps are calculated in terms of volume percentage overlap [82] and Dice coefficient [184]. We denote the segmentation of observer $O1$ and $O2$ for a single region i (midbrain, SN left or SN right) as $S_{i,O1}$ and $S_{i,O2}$. Then, the coefficients for region i are calculated as:

$$\text{Vol \% overlap}_i = \frac{S_{i,O1} \cap S_{i,O2}}{S_{i,O1}} \tag{4.2}$$

$$\text{Dice}_i = \frac{2\,|S_{i,O1} \cap S_{i,O2}|}{|S_{i,O1}| + |S_{i,O2}|} \tag{4.3}$$

Correlation coefficients for volume measurements:
Apart from volume overlaps, we report measures for inter-observer correlation in form of Pearson's correlation coefficient and intra-class-correlation coefficient (ICC) according to Shrout and Fleiss [145].

Pearson's correlation coefficient, or Pearson's r, is a measure for linear inter-dependence of two random variables X and Y. Under the assumption of a normal distribution for both variables and for large sample sizes, r is unbiased and efficient. However, it is defined for any bivariate probability distribution, as long as the covariance of the joint population can be calculated and the variances of the individual populations are non-zero. Hence, Pearson's r can be applied for examination of correlation across both observers in our study. It can take the values $[-1, \ldots, 1]$, with $+1$ indicating perfect positive correlation (i.e.$X = Y$) and -1 indicating perfect negative linear correlation (i.e.$X = -Y$). With X_i denoting the i-th measurement of X and $\bar{X}$ denoting the sample mean of X, Pearson's r is defined as [131]:

$$r = \frac{\sum (X_i - \bar{X})\,(Y_i - \bar{Y})}{\left[\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2\right]^{1/2}} \tag{4.4}$$

In our case, X would represent ratings of observer $O1$ (e.g. volumetric midbrain or SN segmentations) and Y would represent the ratings of $O2$, assuming a normal distribution for volumetric measurements of both raters. Both X and Y are mean-normalized in the calculation of Pearson's r. Results are given in section 4.3.4.

The intra-class-correlation coefficient (ICC) is a measure which is often used for inter-observer reliability measurements in clinical settings and was also used for a four-rater inter-reliability study of SN area measurements in 2D TCUS for Parkinson's Disease diagnosis [99].

In general, the ICC correlation coefficient gives an estimate for the measure of agreement between observers (also called raters, judges, experts, instruments, sensors, or in our case, physicians) who give ratings for targets (also called values, scores or measurements, i.e. in our case, area and volumetric measures for midbrain and SN hyper-echogenicities). The measurements are assumed

to be parametric, i.e. continuous under a normal distribution. Compared to Pearson's r, the ICC coefficient takes the scale of measurements into account. Also, it can summarize and express the inter-rater correlation and reliability across more than two raters in one value.

Shrout and Fleiss explain six different versions of the ICC correlation coefficient [145], denoted as $ICC(m,k)$, where m stands for the ICC model assumption, with $m = [1,2,3]$, and k denotes the number of ratings per observer, where k can take the values $[single,\ mean]$.

Here, we describe the explanations for different model assumptions m as described by Shrout and Fleiss [145]:

- **Model** $m = 1$**:** In this model, it is assumed that raters are randomly chosen and form a subset of the entire set of possible raters. Furthermore, measurements are obtained by different raters from this subset. In our setting, this model would be applicable if each study subject is scanned and segmented by any individual rater randomly selected from a larger set of available expert raters. For example, this applies if different physicians at different clinics segment 3D-TCUS volumes acquired at their respective clinic and their ratings are merged at our site for analysis.
- **Model** $m = 2$**:** In this model, it is assumed that each rater gives a measurement for all targets in the study set. In our setting, this would mean that every physician gives a segmentation for every study subject.
- **Model** $m = 3$**:** In this model, no assumptions about the methods or raters are made.

Furthermore, the explanation for the parameter $k = [single,\ mean]$ can be given as:

- **single** $(k = 1)$**:** The "single-rating" assumption is applicable if each rater gives only one measurement each for the measured target. In our case, this would be applicable if the expert physician segmented the 3D-TCUS volume only once and the obtained volumetric values were used directly for ICC calculation.
- **mean** $(k > 1)$**:** The "mean-rating" assumption is applicable if each rater gives several measurements for each measured target, or if a group of raters forms a consensus opinion about the target to be measured. In our case, this would be applicable if the expert physician segmented each 3D-TCUS volume several times and e.g. the average volumetric values were used for ICC calculation. As another scenario, the "mean-rating" assumption would be applicable if e.g. several physicians from one hospital each segmented a respective volume and the average volume was correlated against the average volume values obtained from several raters of another hospital.

Given our dataset and segmentation setting, we model the observers $O1$ and $O2$ as random effects, where each observer segments every single 3D-TCUS volume from our dataset. Furthermore, we correlate one single rating per

target, i.e. one segmentation per observer for each 3D-TCUS volume. Hence, we could set $m = 2$ and $k = 1$ and use $ICC(2,1)$ for calculation of correlation. According to Shrout and Fleiss [145], the ranges of reliability can be assessed as: 0.00-0.10, virtually none; 0.11-0.40, slight; 0.41-0.60, fair; 0.61-0.80, moderate; 0.81-1.00, substantial.

In terms of implementation, we use the programming language and statistical data analysis tool R [126]. The *ICC* methods are implemented in the R library package *"psych: Procedures for Psychological, Psychometric, and Personality Research"*, implemented by William Revelle [130].

4.3 Results of SVM-guided PD classification and multi-observer study

In the following paragraphs, we report our results on the multi-observer study concerning volumetric measurements of midbrain and SN, single-dimensional diagnostic classification using ROC analysis, multi-dimensional diagnostic classification using non-linear SVM classification and a first impression of the inter-observer variability of the 3D method, given our two-person observer study.

4.3.1 Midbrain and SN echogenicity volume measures

Tables 4.1 and 4.2 give an overview of patient details such as subject class (HC = healthy control, PD = Parkinson patient), sex (F = female, M = male) and age. Quantitative values measured by observers one and two are given in form of value pairs O1/O2 for volumes of midbrain, SN left and SN right as well as the largest 2D slice area of SN echogenicities. For PD patients, we additionally list information about the Hoehn & Yahr stage, the diagnosed dominant side, the duration of the disease and the diagnosed Parkinson type (AR = akinetic rigid, TD = tremor dominant, M = mix-type).

Healthy Controls	**Mean**	**Std.Dev.**	**Min**	**Max**
Sex	6 female / 5 male		-	-
Age	55.6	9	43	73
Vol. Midbrain [mm3]	492.8 / 444,5	72.3 / 54.36	396.5 / 356.4	606.1 / 531.1
Vol. SN Left [mm3]	62.2 / 57.9	42.6 / 29.7	0 / 3.1	122.8 / 102.6
Vol. SN Right [mm3]	72 / 66.6	51.2 / 28.3	11.0 / 21.8	176.6 / 106.2
Max. Area 2D [mm2]	25.6 / 22.5	12.6 / 7.6	7.3 / 11.1	41.7 / 37.7

Table 4.1: Details for subjects from the Healthy Control (HC) group. Volume and area values are given for observer one and two (reported as value pairs O1/O2).

4.3.2 Single-dimensional Classification with ROC Analysis

Using the receiver-operating characteristic (ROC) analysis for single-dimensional feature classification, we calculated the threshold or cut-off value which yielded the highest f-measure of all points with a classification result above the diagonal in ROC space (cf. table 4.3). Here, we report the classification results for each of the five single-dimensional features segmented by observers one and two. The area-under-the-curve (AUC) value furthermore provides a measure for the overall quality of the ROC curve, in particular if classification results in terms of sensitivity, specificity and f-measure are identical for several classifications (e.g. for observer two).

The best classification result was obtained for the features ”Larger side 3D SN volume (absolute)” for both observers O1/O2, reaching a sensitivity

Parkinson Patients	Mean	Std.Dev.	Min	Max
Sex	4 female / 7 male		-	-
PD Dominant Side	4 right / 7 left		-	-
PD Type	4 AR / 3 TD / 4 M		-	-
PD Hoehn&Yahr	2.1	0.8	1	3.5
PD Duration [years]	7.4	3.8	3	16
Age [years]	65.8	7.9	45	73
Vol. Midbrain [mm3]	572.6 / 487.5	95.0 / 79.4	379.6 / 317.2	736.7 / 597.1
Vol. SN Left [mm3]	108.5 / 117.5	56.2 / 64.7	24.1 / 35.0	213.1 / 231.5
Vol. SN Right [mm3]	103.4 / 91.8	38.9 / 22.1	33.6 / 63.6	161.8 / 124.0
Max. Area 2D [mm2]	35.2 / 32.1	9.0 / 12.2	20.3 / 19.24	49.8 / 55.9

Table 4.2: Details for subjects from the Parkinson Disease (PD) patient group. Volume and area values are reported as value pairs for observer one and two (O1/O2) (Abbreviations for PD Type: AR = akinetic rigid, TD = tremor dominant, M = mix-type).

of 90.9/100.0%, specificity of 63.6/54.6%, f-measure of 0.8/0.81 and AUC of 0.78/0.84, for the volume cutoffs or thresholds 89.4/76.4mm3. The feature "Sum of 3D SN volumes (absolute)" yielded similar classification results, while relative volume values as well as the feature "Largest 2D area" performed notably worse in terms of classification accuracy. Figure 4.2 shows the ROC curve for observer $O1$ on the best-performing single-dimensional feature "Larger side 3D SN volume (absolute)".

Feature	Sensitivity	Specificity	f-Measure	AUC	Threshold
Larger side 3D SN volume (absolute)	90.9/100.0%	63.6/54.6%	0.80/0.81	0.78/0.84	89.4/76.4 mm3
Larger side 3D SN volume (relative)	63.6/100.0%	72.7/54.6%	0.67/0.81	0.69/0.78	0.44/0.31 Vol%
Sum of 3D SN volumes (absolute)	100.0/100.0%	54.6/54.6%	0.81/0.81	0.78/0.81	138.7/152.5 mm3
Sum of 3D SN volumes (relative)	72.7/100.0%	63.6/54.6%	0.70/0.81	0.69/0.75	0.63/0.50 Vol%
Largest 2D area	81.8/72.7%	54.6/54.6%	0.72/0.67	0.69/0.74	30.4/23.5 mm2

Table 4.3: Classification results with a single-dimensional feature and a single threshold chosen from a ROC analysis of segmentations from observers one and two (results reported as value pairs O1/O2).

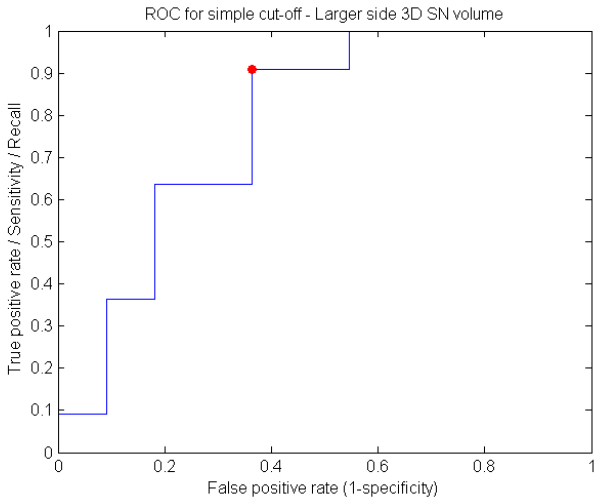

Figure 4.2: ROC curve plot for observer $O1$ on the best-performing single-dimensional feature "Larger side 3D SN volume (absolute)" ($AUC = 0.78$). The threshold with the best F_1-score is marked with a red dot, at $\tau = 89.4mm^3$.

4.3.3 SVM Classification on Multi-Dimensional Features

Performing a SVM classification using multi-dimensional features, the best results for both observers O1/O2 were achieved by using the three volumes "Volume SN Left, Volume SN Right and Volume Midbrain (absolute)", which yielded 90.9/72.7% sensitivity and 72.7/81.8% specificity at f-measures 0.83/0.76. If only the two features "Volume SN Left and Volume SN Right (absolute)" are used, the sensitivity decreases significantly to 63.6/72.7%, while the specificity remains equal at 72.7/72.7%. Incorporating the midbrain volume by using it as the normalizing factor in the feature set "Volume SN Left and Volume SN Right (relative)", both sensitivity and specificity are reduced to 72.7/72.7% and 63.6/63.6%, respectively. In relation to the 2D method, we also performed an SVM classification using the single-dimensional feature "Largest 2D area", which yielded the same classification performance as in the ROC analysis for observer O1 (sensitivity 81.8%, specificity 54.6%). For observer O2, a significantly higher sensitivity but lower specificity of 90.9% and 45.5% respectively is achieved. A summary of classification results with SVM on single- and multi-dimensional features is given in Table 4.4. Two further observations shall be noted here. First, including the midbrain volume as a separate dimension in the feature set "Volume SN Left, Volume SN Right and Volume Midbrain (absolute)" yielded the best result, while incorporating it as a "normalizing factor" by dividing the SN volumes with the midbrain volume in the feature set "Volume SN Left and Volume SN Right (relative)" yielded notably worse results. Second, the second rater $O2$ consistently showed higher sensitivity and lower specificity than rater $O1$. We will comment on both these observations and give possible explanations in the discussion section.

Feature	Sensitivity	Specificity	f-Measure
Single-dimensional Features			
Largest 2D area	81.8%/90.9%	54.6%/45.5%	0.72/0.74
Multi-dimensional Features			
VolumeSNLeft + VolumeSNRight (absolute)	63.6%/72.7%	72.7%/72.7%	0.67/0.73
VolumeSNLeft + VolumeSNRight (relative)	72.7%/72.7%	63.6%/63.6%	0.70/0.70
VolumeSNLeft + VolumeSNRight + VolumeMidbrain (absolute)	90.9%/72.7%	72.7%/81.8%	0.83/0.76

Table 4.4: SVM classification results on single- and multi-dimensional features extracted from the expert 3D US volume segmentations of observers one and two (reported as value pairs O1/O2).

4.3.4 Inter-observer Comparison

For inter-observer correlation, we report volume segmentation overlaps (Dice and volume percentage overlap) as well as correlation of volumetric measures across both observers (Pearson and ICC correlation coefficients). The volume overlaps for segmentation of midbrain are fairly consistent (Dice 0.85), while SN overlaps are considerably smaller (Dice ~ 0.55), which can be partly accounted to the small and patchy spatial distribution of SN echogenicities. Hence, it is important to note that despite seemingly low overlap of SN segmentation, the Pearson correlation was moderate to substantial at 0.77-0.82, for all three 3D values. Intra-class correlation ICC(2,1) was fair (0.60) for the midbrain region volume measures and moderate for SN right (0.69) and SN left (0.78). All Pearson and ICC correlation coefficients of volumetric measures in 3D were significant at $p < 0.001$. Correlation for the maximum 2D area across both observers was substantially lower, i.e. Pearson correlation of 0.54 and ICC(2,1) of 0.50, both significant at $p < 0.01$. A summary of multi-observer study results is given in table 4.5.

Coefficient	Midbrain	SN Right	SN Left	Max Area 2D
Dice volume overlap	0.85	0.57	0.54	–
Volume percentage overlap	79.8%	58.7%	56.9%	–
Volume Correlation (Pearson)	0.82***	0.80***	0.77***	0.54**
ICC2 (single, random raters)	0.60***	0.69***	0.78***	0.50**
	Notes: * significant at $p < .05$; ** $p < .01$; ***$p < 0.001$			

Table 4.5: Results of inter-observer analysis between observer one and two, reported as volume overlap values (Dice and volume percentage) and feature correlation (Pearson and ICC).

4.4 Discussion and Future Work

In this chapter, we studied whether using volumes of the midbrain and signal alterations of the SN might achieve similar or possibly better discrimination between healthy and diseased subjects as compared to using 2D areas in the traditional method (see chapter 1.3). We presented results from two observers in order to provide an initial estimate of inter-observer variability and reproducibility of the method.

Our approach showed a low rate of drop-outs (1/23) due to no or partial visualization of intracranial structures whereas normally drop-outs range between 10-20% [167]. One of the reasons could be that our technique includes a bilateral reconstruction which helps balancing different acoustic bone windows in one subject. A limit of our study is a relatively small number of subjects studied, which we consider adequate for a pilot study however.

It should be noted that although we used two observers for segmentation of 3D B-mode volumes, we used only one examiner during 3D-TCUS scanning of subjects. Although this prohibits the assessment of objectivity and repeatability during the stage of 3D acquisition, we argue that scanning in 3D implicitly increases objectivity and repeatability, since the selection of an optimal cut-plane becomes irrelevant. Moreover, several previous studies have shown that assessment of anatomy using 3D freehand ultrasound scans compared to 2D leads to results which are reproducible, increase objectivity, diagnostic relevance and repeatability of results (also longitudinally), e.g. in 3D-US of the carotid artery [6], of fetal liver volume [36] or prostate volume [155].

For the expert observer, the results of our pilot study may indicate a slight advantage of the 3-dimensional method as compared to conventional area measurements for differentiating patients and healthy control subjects. In relation to previous blinded studies in 2D, the sensitivity turned out to be comparable or even slightly better in our study, while the specificity is slightly lower [58, 124]. It has to be mentioned that the 2-dimensional measurement was applied after reconstruction of the data and might therefore not be directly comparable to the so far used 2-dimensional real-time technique.

As mentioned in the results section, it is interesting to note that including the midbrain volume as a normalizing factor seems to improve the classification results in the multi-dimensional classification case, however only if it is included as a separate feature dimension, i.e. in the feature set "Volume SN Left, Volume SN Right and Volume Midbrain (absolute)". Compared to the self-defined normalization by division in the feature set "Volume SN Left and Volume SN Right (relative)", both sensitivity and specificity increase notably. These results seem to indicate that it is preferable to let the classifier (e.g. SVMs) select optimally separating hyperplanes between classes, rather than manually defining intuitive (yet somewhat arbitrary) inter-feature operations such as dividing SN volumes by the midbrain volume in order to obtain relative volumes.

The second, novice observer achieved comparable classification results for both single-dimensional features (cf. ROC analysis, table 4.3) and multi-dimensional features (cf. SVM classification, 4.4). In single-dimensional classi-

fication, the largest AUC value was obtained for the feature "Larger side 3D SN volume (absolute)", which was also one of the best classification features for observer 1. It is notable that despite different experience levels with the TCUS technique, both observers had better classification results when 3D volumetric features were used for classification, compared to the largest 2D area. Multi-dimensional classification using SVMs and 3D volumetric features only yielded noticeably better results for observer $O1$, and the best feature set for both observers was a combination of absolute volume measures of midbrain, SN left and SN right.

Again, as mentioned in the results section, it is noteworthy that the specificity of the novice observer $O2$ was consistently lower and sensitivity consistently higher than for $O1$, i.e. the expert. Both aspects indicate an oversegmentation of SN hyper-echogenicities in the case of $O2$. There are several possible explanations for this. First, the expert observer $O1$ had more experience with the 2D transcranial ultrasound method and the medical background, probably allowing for more careful and specific selection of hyper-echogenic speckle patches in the SN region. In contrast, $O2$ had more experience with reconstructed 3D freehand ultrasound volumes and their appearance in general. Also, $O2$ had more experience and practice with the segmentation tool, ITK-Snap **??**, probably allowing for finer segmentation within the same timeframe as $O1$. It has been shown that the user-friendliness of manual segmentation tools can produce noticeable effects on segmentation outcome and potentially introduce systematic errors or biasses on the segmented regions

Despite the mentioned differences, when inter-observer correlation, one can note that while volume overlap coefficient e.g. in terms of Dice coefficient was lower for SN regions than for the midbrain, volume measurements correlated comparably for all three regions across the two observers. Volumetric measurements correlated better than the maximum 2D area of SN echogenicities.

At this point, we would like to point out a study by van de Loo et al. [99], who performed a more extensive intra- and inter-rater observability study on the 2D TCUS method for Parkinson diagnosis. They compared 2D area measurements of four independent and fully blinded raters on 32 subjects (22 PD, 10 healthy controls). They achieved substantial intra-rater correlation with ICC(3,k) values of up to 0.93 and 0.97 for both hemispheres and substantial inter-rater reliability with ICC(2,k) of 0.84 and 0.89 across raters. It is important to note that these excellent result probably result from the fact that all four judges were highly experienced expert raters concerning the 2D TCUS methods. One should also note that other groups have had difficulties reproducing such excellent rating results with the 2D method and argue that it is difficult to reproduce them, without the substantial experience of the reference research groups [162]. At this point, we again would like to note that in this pilot study, we had considerably higher correlation across raters given 3D volumetric measures, rather than 2D measures. Although these are very early results and have to be confirmed by other groups researching the TCUS method in 3D in future, they are complementary with the above mentioned observations from groups with varying experience levels.

Several independent studies on more than 100 subjects have shown that 90% of PD patients show an increased SN echogenicity (i.e. sensitivity of 90%) [20]. At the same time, these studies showed that only around 10% of non-PD subjects present an enlarged SN echogenic area, which corresponds to a specificity on the order of 90%. Mostly, the reduced specificity in our study was due to four of the healthy subjects (36.4%) showing an unusually large SN echogenicity, also in 2D, i.e. before 3D reconstruction. This might be due to the THI imaging modality chosen, which reportedly over-estimates SN echogenicity (personal communications) but has also shown more consistency in the past [125]. Since we applied this modality in all subjects, however, we expected this error to be systematic and compensable by non-linear classification of SVMs. Another possibility could have been a much better pre-auricular acoustic bone window with clearer echo-responses, possibly due to the younger average age of the healthy control group. Our future studies will investigate this notion by introducing a more rigorous age matching scheme. Another possibility is that these larger SN echogenicity are already a marker for imminent PD onset, which is a possible explanation according to [24], who demonstrated on more than 1800 subjects that enlarged SN echogenicity in healthy controls correlates with a relative risk of about 17 and therefore seems to serve as a marker for early detection of PD.

Given the presented results, 3D-TCUS has a high potential for PD diagnosis due to the higher amount of information contained in the 3D volume as compared to a 2D image. Please note that while no real-time 2D echography was used for comparison, our feature "Largest 2D Area" provides an analogous comparison to the 2D method. Thus a comparison with the "gold standard" method of the so far used 2D method seems reasonable.

We would like to clearly point out that our study was designed as a first feasibility study of 3D TC-US for Parkinson diagnosis, with a first quantitative evaluation. The results we present, in particular the concrete values e.g. for thresholds of SN volumes, have to be interpreted with care. The reference groups for the 2D method suggest a much larger population size for determining a reasonable cut-off value of 2D SN echogenicity area. We are planning such a bigger cohort examination in a future study (see chapter 8.2.2), which may also allow differentiating between the different subtypes of PD concerning their specific volumetric and spatial alterations. The values and classification results presented in this chapter, however, can serve as a first basis for discussion with other research groups who plan to perform 3D TC-US.

4.5 Conclusion

Hyper-echogenic signals from the midbrain of patients with Parkinson's Disease can be recorded and measured with a 3D Freehand US technique as presented in the previous chapters. The first results as presented in this chapter are promising. Compared to previously used 2D data acquisition and evaluation, the usage of 3D acquisition and data analysis may allow for a more objective and less user-dependent diagnosis of Parkinson's disease and other pathologic brain alterations.

However, although it was necessary to first create a manual segmentation by human experts in order to lay the groundwork for this thesis and a gold-standard for our dataset, average durations for manual segmentation of 20 minutes per dataset clearly show that the 3D method imposes an overhead on the physician in terms of data processing. Despite advantages of the 3D method, this could potentially hinder an adoption of the technique into clinical practice and diagnostic routine. Also, there is still a non-negligible deviation left between human rater opinions, in terms of segmentation overlap and volumetric measures.

In the following chapters, we will thus present our research on various methods for automatic and computer-assisted segmentation methods for midbrain (chapter 5) and SN hyper-echogenicities (chapter 6), which significantly reduce data processing and user interaction time for the physician, while potentially further increasing the objectivity of the volumetric labeling.

5

Midbrain Segmentation in 3D Transcranial Ultrasound Data

In the introduction , we have learned that 2D-TCUS is a cheap, quick and completely non-invasive method for diagnosis of Parkinson's Disease [14, 168, 22], even before onset of motor symptoms [24]. This makes the TCUS method a potential technique for PD screening and early-detection of future patients. However, we have also learned that the technique suffers from certain shortcomings, most of all the difficulty of the sonographic acquisition, which leads to a high intra- and inter-rater variability. This largely depends on the experience of the sonographer [162].

In the last chapter, we have motivated the acquisition of TCUS in 3D, arguing that 3D-TCUS makes this method easier and more objective, e.g. by rendering the difficult selection of an optimal cut-plane obsolete. Our experiments in a two-person inter-observer variability study give a first indication that indeed, 3D volumetric features (both stand-alone, i.e. as single-dimensional features, as well as in multi-dimensional combination of features) lead to improved classification of healthy controls versus PD patients, compared to classic 2D area features.

However, our experiments have also clearly demonstrated the additional data processing overhead in 3D. Given that manual segmentations in 3D take an average of 20 minutes per subject, the applicability of the 3D method in clinical routine becomes questionable. Also, segmentations in 3D still leave a non-negligible margin of inter-observer variability.

In this and the following chapter, we are introducing techniques with which we intend to make the 3D method more clinically practical by introducing computer-aided, (semi-)automatic segmentation and voxel classification methods, first through (semi-)automatic segmentation of the midbrain and then through automatic segmentation of substantia nigra (SN) hyper-echogenicities (see chapter 6). This reduces the data processing time to clinically acceptable

durations around one minute. Most importantly, however, it allows for a more systematic processing of data, which could establish a more objective segmentation outcome and more comparable results across different clinical groups analyzing the 3D data. Ultimately, the aim is to create methods which allow for systematic, objective and diagnostically accurate computer-aided data processing in 3D across different clinical centers.

5.1 Objective and Overall Approach

5.1.1 Motivation for Midbrain Segmentation in 3D-TCUS

The crucial part of any computer aided 3D-TCUS diagnosis system is an accurate segmentation of the whole midbrain. In fact, it is advisable to segment the midbrain first and the SN afterwards, because SN hyper-echogenicities appear to be almost identical to speckle and intensity patterns outside the midbrain area (cf. US images in figure 5.1). This was also proven by related work on computer aided PD diagnosis, where either a manual [87, 37] or an automatic [52] midbrain segmentation is used to confine the search space for SN hyper-echogenicities. As a consequence, any method which provides a robust and reliable midbrain segmentation from transcranial US greatly facilitates the subsequent steps of SN segmentation and classification. Please note that the only related approach proving an automated midbrain segmentation is the one of [52]. This method is, however, restricted to 2D and based on a complex and computationally intensive finite-element model. We are not aware of further attempts at computer-assisted segmentation of transcranial B-Mode US. We argue that this is likely because of the difficulties in defining appropriate prior knowledge, data term and optimization methods (see section 5.4).

5.1.2 Overall Approach

The automatic segmentation of the midbrain in 2D or 3D-TCUS is a relatively difficult task. Due to the scanning at low frequencies (2-4 MHz), transcranial US is unfortunately poor in resolution and characterized by high levels of noise and large speckle patterns (see Fig. 5.1). Also, the US acquisition through the temporal bone window, a layer of bone thin enough to be penetrated by low-frequency US (typically 2-4MHz), introduces additional phase aberrations and interferences [75] and higher absorption than in regular tissue, causing low contrast.

Due to these difficulties, we first applied several robust segmentation methods developed at the CAMP group on our 22-subject data. The methods included an approach based on a hybrid deformable model using non-uniform rational B-splines (NURBS) [50] as well as a variational approach using polar active contours [10, 11, 12]. Unfortunately, both methods failed in properly segmenting the midbrain area, despite their designs towards robustness in face of low-contrast boundaries, and low tendency towards leakage. Failure occured partly due to significant leakage through gaps in the midbrain boundary and partly due to anatomically incorrect shapes of final segmentation results.

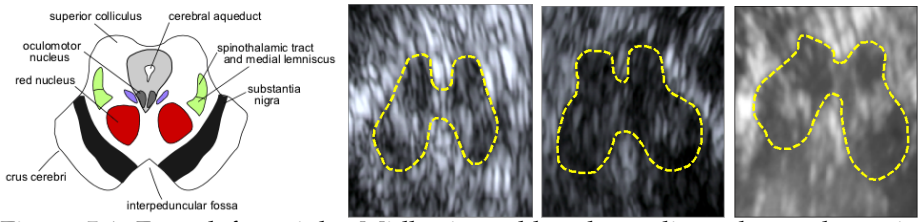

Figure 5.1: From left to right: Midbrain and basal ganglia such as substantia nigra, along with typical 2D TC-US images (middle), and a slice through one of our 3D volumes.

Consequently, we developed a new approach, for which we had to combine several components which allow for explicit 3D surface segmentation in 3D-TCUS. The design goals were high anatomical accuracy of the final result as well as high robustness, despite missing boundaries, low overall contrast and high variability across subjects due to different bone windows. The final approach combines an explicit Active Surface Model (ASM) based on the "Active Polyhedron" method [151] with a region-based cost-function adapted to US through localized statistics [95]. Most importantly, we incorporated a statistical shape model for anatomical shape prior and surface regularization based on spherical harmonics [138]. Lastly, we developed and implemented a gradient-descent optimization method.

The overall method based on a US-adapted Active Surface Model (ASM) was published at MICCAI 2011 [2]. In the following section, we will explain the necessary theoretical background on statistical shape models (SSM) and active shape model segmentation (ASM). Subsequently, we will explain our proposed segmentation method in detail, describe the performed experiments and report the results on our 22-subject dataset.

5.2 Statistical Shape Modeling Background

Object recognition and image segmentation can be particularly difficult in ultrasound data, due to attenuation, speckle and artifacts like shadowing. This can lead to low-contrast regions and gaps in anatomical boundaries, which cause edge cues and region-based information to be often insufficient for reliable and accurate segmentation [110].

In the mid-nineties, several works were published on statistical modeling of shape and how to incorporate a learned shape model into previously proposed segmentation techniques such as snakes and active contours [85]. The incorporation lead to a dramatic increase in segmentation performance and robustness, even under very difficult conditions such as high noise and low-contrast partial boundaries, as demonstrated for example on echocardiography images in the seminal work by Taylor et al. [41]. Since its first appearances in segmentation literature, the application of learned statistical models of shape has had a profound impact in the community and found many applications. In 2011, for example, around one fifth of all proceedings papers at the MICCAI conference featured the search terms "active shape" or "active appearance".

In the midbrain segmentation approach in our work, statistical shape models (SSM) play an important role for high robustness and segmentation quality as well. In the following, we give a brief overview over the theoretical background and some selected approaches in literature.

5.2.1 Shape model construction

Statistical shape modeling (SSM) describes the process of learning the shape of a desired target anatomy and its natural shape variations in a training stage. In a later application stage, e.g. for segmentation of the anatomy in a previously unseen image or image volume, the model can be applied to find and segment the target anatomy even under difficult circumstances such as high noise or partial occlusion.

The first step of SSM design is to select a model representation. The choice is whether to use an explicit model representation (also called finite-dimensional SSM [68]) or an implicit shape representation (also called infinite-dimensional SSM [68]). Before proceeding, we would like to point out that we utilize an explicit shape and SSM model representation, since our overall segmentation method is based on an explicit active-contours approach with shape prior. The alternative would be to use an implicit or infinite-dimensional representation using level-sets and statistical deformation models (SDM). In this thesis, we will only explain the explicit SSM modeling workflow. For more information on existing techniques for implicit, level-set based SSM modeling, we refer the reader to the recent literature review by Cramers et al. [43].

The basic principle of explicit SSM model creation is to sample M D-dimensional training shapes with the same number K of vertices, where each vertex $v_k = (x_k, y_k, z_k)^T$ corresponds to the same location or shape feature on all training shapes. These salient shape points are often referred to as "landmarks" [41, 68], however despite the term, these landmarks do not necessarily have

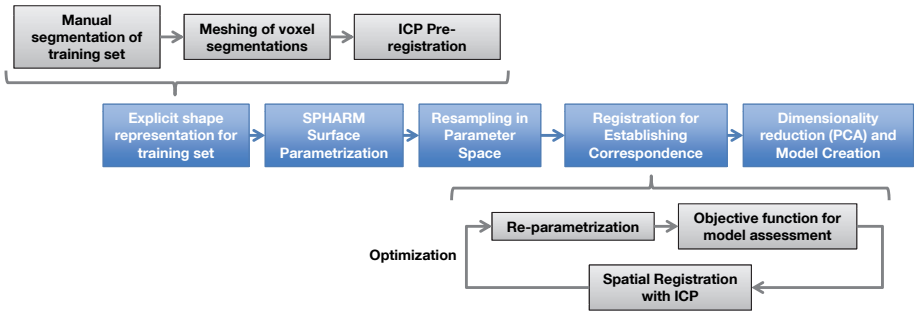

Figure 5.2: Workflow steps for SSM model building in this thesis, based on Spherical Harmonics (SPHARM) surface parameterization [138].

to coincide with anatomical "landmarks" defined by physicians, e.g. for landmark-based registration. If the M training shapes are spatially registered, each of the K landmarks has corresponding sample points on the M training surfaces, which form a point cloud. In SSM literature, this point cloud is referred to as a "Point Distribution Model" (PDM) [41]. By concatenating the D dimensions of shape vertices into one long vector (the so-called "shape vector" [41, 40]), we obtain M samples in a $(K \cdot D)$-dimensional space. A SSM model is then obtained by applying a linear or non-linear dimensionality reduction technique [48] to this sample distribution, which is able to approximate the $(K \cdot D)$-dimensional manifold of legal shapes with a linear or non-linear model. This model can then be used generatively, i.e. arbitrary in-between shapes can be generated by the model.

Figure 5.2 shows a more detailed workflow of steps performed for SSM model building in this thesis, using *"SPHARM-MAT"*, a MATLAB toolbox for SSM creation based on a Spherical Harmonics (SPHARM) surface parameterization [138]. While it would be exhaustive to explain each step in full detail (see the excellent book on SSM theory and application by Davies et al. [48] for a more thorough coverage of all following aspects), the following section will briefly explain the most important concepts for the five main steps for SSM generation using the *"SPHARM-MAT"* toolbox. Alternative approaches used in literature will be briefly touched in chapter 5.2.1.6.

5.2.1.1 Explicit Shape Representation of Training Set

The first step of SSM generation is to assemble a set of M training shapes. For explicit SSM modeling, the training shapes have to be established in form of explicit meshes, in most cases triangular. There are several ways to accomplish this. The first way is by manual placement of K vertices on the training images, at salient shape features such as corners or dents. In practice however, when dealing with medical images, this approach is not feasible. This is true in particular for 3D surfaces of human anatomy, which are often sampled with several hundred vertices, partly placed in surface regions of smoothly shaped anatomy without salient features, making a consistent and corresponding manual landmark placement across M training shapes impossible.

Instead, SSM modeling of human anatomies typically begins with M manual segmentations which are provided as binary voxel volumes, e.g. of a single organ such as the liver, or as in our case of the midbrain (see workflow, figure 5.2, top row). Before SSM creation, the voxel volumes of training segmentations have to be transformed to explicit meshes. One straightforward way would be to assemble an outer surface mesh by sub-dividing each voxel outer surface into two triangles, however this creates very fine-grained meshes.

In our work, we utilize a meshing tool called *"iso2mesh"* [53], which is a free and open-source MATLAB-based toolbox for mesh generation and processing. Its features include a method called `vol2surf`, for converting a binary (or multi-valued) image segmentation volume to one (or several) triangular surface meshes. Compared to pure voxel-surfaces, which could be obtained using e.g. a simple approach such as marching-cubes segmentation, the surface created by `vol2surf` is sub-sampled and smoothed, which can later increase the speed and robustness of further SSM creation.

The conversion is based on a restricted version of the original Delaunay triangulation [49, 25]. The voxel volume is first represented as an implicit function, where the zero-level-set denotes the object's outer surface. Boundary points are sampled from this zero-level-set and triangulated three-dimensionally. The size of the sample set is iteratively increased, as in a Delaunay refinement process, until user-specified size and shape criteria of triangular surface elements are satisfied. The criteria are 1) an angular bound, i.e. a lower bound in degrees for the angles of mesh facets, 2) a radius bound, i.e. an upper bound on the radii of surface Delaunay balls and 3) a distance bound, i.e. an upper bound for the distance between the circumcenter of a mesh facet and the center of a surface Delaunay ball of this facet. The implementation is provided by a wrapper function to "cgalsurf", which itself is a sub-feature of the *"Computational Geometry Algorithms Library"* (CGAL) [1], a powerful C++ library for efficient and reliable geometric computation. We constrained the resolution of the surface mesh generated by `vol2surf` by setting the maximum radius of the Delaunay sphere to 2mm for the midbrain surface.

One final step is to pre-register the triangulated surfaces, e.g. using Iterative-Closest-Points registration (ICP) [26] or Soft-Assign Procrustes Matching (SPM) [127]. Although the process of SSM model creation incorporates an iterative update of surface registrations for establishing correspondence (see chapter 5.2.1.4), we found it advisable to perform a pre-registration, partly using ICP alone, partly even using a manual pre-alignment plus ICP, in particular if body axes were too largely mis-aligned across 3DUS volumes to be recovered by ICP. As a reference surface for ICP, we selected the shape with median volume and performed $(M-1)$ pair-wise registrations.

5.2.1.2 Surface Parameterization

Surface parameterization is a bijective (i.e. one-to-one) mapping between the object space, where vertices are expressed in form of cartesian coordinates $v_k = (x_k, y_k, z_k)^T$, and a parameter space which often has preferable attributes over a cartesian representation.

For shapes belonging to the class of closed 2D-manifolds of genus zero, i.e. closed-shape surfaces without holes, the base domain is the unit sphere. The parameterization on the unit sphere is represented by spherical coordinates (θ, ϕ), with $\theta \in [0, \pi]$ taken as the polar (colatitudinal) coordinate and $\phi \in [0, 2\pi)$ as the azimuthal (longitudinal) coordinate [141]. Representing shapes with spherical topology using polar parameterization has preferrable properties, as we will see later.

Following the notation of Davies et al. [48], we denote the parameterization as a continuous, one-to-one mapping $\mathcal{X}_i$ from the spherical parameter space X to the shape S_i as:

$$X \overset{\mathcal{X}_i}{\longmapsto} S_i, \ \mathbf{x} \overset{\mathcal{X}_i}{\longmapsto} \mathbf{S}_i(\mathbf{x}), \tag{5.1}$$

where $\mathbf{x} \in X$ is a vector-valued parameter vector (e.g. (θ, ϕ)), and $\mathbf{S}_i(\cdot)$ is the vector-valued shape function associated with shape S_i. Using this parameterization, we can now express vertex coordinates $v = (x, y, z)^T$ as three distinct functions $v(\theta, \phi) = (x(\theta, \phi), y(\theta, \phi), z(\theta, \phi))^T$.

In the case of the *"SPHARM-MAT"* toolbox, the parameter space of polar coordinates is further expanded into another parameterization based on Spherical Harmonics (SPHARM), which are an extension of Fourier techniques to 3D [139]. Like most other parameterization methods used in SSM literature [68], Spherical Harmonics (SPHARM) [29] parameterization is limited to shapes of spherical topology.

Using SPHARM [29], the already parameterized surface vertices $v(\theta, \phi)$ can be further parameterized using an expansion of the polar coordinate space to Fourier spherical harmonic basis functions.The expansion can be denoted as [139]:

$$v(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_l^m Y_l^m(\theta, \phi), \tag{5.2}$$

where $Y_l^m(\theta, \phi)$ are the spherical harmonic basis functions of degree l and order m:

$$Y_l^m(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\phi}, \tag{5.3}$$

and $P_l^m(\cos\theta)$ are the associated Legendre polynomials defined by the differential equation:

$$P_l^m(x) = \frac{(-1)^m}{2^l l!} (1 + x^2)^{\frac{m}{2}} \frac{d^{l+m}}{dx^{l+m}} (x^2 - 1)^l. \tag{5.4}$$

Similar to a representation using Fourier basis function, a 3D surface can be approximated with increasing accuracy if the degree of spherical harmonics is increased, i.e. with increasing degrees l up to a user-specified maximum degree L_{max}, which we set to 15 in our work. The value 15 is the maximum suggested value in the toolbox. We choose it because first, the usage of less spherical basis functions does not introduce a noticeable computational advantage during segmentation, and second, we already introduced a smoothing of training shapes due to the usage of the `vol2surf` surface extraction method (see above). By using spherical basis functions up to the degree 15, we thus avoid further

smoothing which would occur due to capping of higher-frequency spherical harmonics. When working with direct voxel surfaces of the binary training segmentations, however, it is suggested to use a smaller degree of spherical harmonics, in order to avoid encoding e.g. high-energy 90-degree voxel edges in the spherical harmonics of higher order.

The spherical parameterization is obtained by solving for the weights c_l^m, such that each training shape can be expressed in form of the spherical harmonic basis vectors Y_l^m. As an initial parameterization, the weights c_l^m can be obtained using standard least squares estimation [139].

In the *"SPHARM-MAT"* toolbox, the initial least-squares mapping into the parameter space is further refined and smoothed by the method "Control of Area and Length Distortion" (CALD) [142], which iteratively alternates between a local minimization of triangular area distortion in parameter space with simultaneous control of the worst length distortion and a global equalization of area distortions for all mesh vertices over the whole sphere [139].

5.2.1.3 Resampling in Parameter Space

It is important to note that after voxel segmentations have been meshed using *"iso2mesh"* (see section 5.2.1.1), each training mesh has a different number of vertices per training shape. In order to create a point-distribution model (PDM), however, each training shape must be sampled with the same number of vertices. In *"SPHARM-MAT"*, this is achieved by projecting a regular tetrahedron which already closely approximates a unit sphere into parameter space. Thus, the surface's parameter space is re-sampled bi-linearly at a near-uniform distribution of vertices, which can be projected back into cartesian space. In our work, we use a regular tetrahedron with 642 vertices created by a level-4 icosahedron subdivision, i.e. each training shape is then represented by 642 vertices as well.

It should be noted that there are methods which sample all training shapes with the same number of vertices in one pass, some of them even establish correspondence at the same time, as we will briefly explain in chapter 5.2.1.6. However, those are often restricted to image modalities with clearly defined anatomy and anatomic target regions which stand out clearly from the background. Both conditions are unfortunately not given in 3DUS, which is why we resort to this resampling strategy described by Shen et al. [139].

5.2.1.4 Establishing Correspondence

Once the surface is sampled with the same number of vertices for all training shapes, it is necessary to associate regions on all shapes such that each model vertex represents the same shape feature in all training shapes. This is often referred to as the "correspondence problem" and regarded as the most challenging and complex step in SSM model building [68, 48]. Only once correspondence has been established, a proper point-distribution-model (PDM) can be formed, which is the prerequisite for calculating a SSM.

As Heimann et al. [68] suggest, establishing correspondence is basically a registration problem. In their review, they summarize different approaches

in literature, including mesh-to-mesh registration, mesh-to-volume registration, volume-to-volume registration, registration of paramaterizations as well as population-based (i.e. group-wise) registration. Here, we focus on the *"SPHARM-MAT"* registration method, which is a registration of parameterizations.

A straightforward way of establishing correspondence would be to move vertices on the shape surfaces directly, and optimize their location such that a quality measure for correspondence across shapes is maximized. Different point-based and feature-based quality measures (or combinations thereof) can be applied, such as a root-mean-square-distance (RMSD) or comparing curvatures between corresponding vertices on two shapes. However, as Davies et al. point out in [48], moving points directly on the shape surface is equal to generating a diffeomorphism of the shape surface onto itself, which has different properties for each shape. Furthermore, sliding points can move them off-surface, which would require subsequent re-projection.

Instead, it is much more elegant and flexible to establish correspondence by modifying the *parameterizations* for each training shape. The concept behind this idea is explained in detail by Davies et al. in [48], and we will summarize the most important points in the following.

Re-parameterization: From eq. 5.1, we recall that each training shape S_i is represented by a parametric shape function $\mathbf{S}_i(\mathbf{x})$. It is defined by the initial bijective mapping $\mathcal{X}_i$, which describes the projection from parameter space X to the shape S_i as:

$$X \overset{\mathcal{X}_i}{\longmapsto} S_i,\ \mathbf{x} \overset{\mathcal{X}_i}{\longmapsto} \mathbf{S}_i(\mathbf{x}). \tag{5.5}$$

In other words, "the mapping $\mathcal{X}_i$ thus associates a parameter value $\mathbf{x}$ to each point on the i^{th} shape, with the coordinates of that point on the shape being the value of that shape function, $\mathbf{S}_i(\cdot)$" [48]. Formally, correspondence between two shapes is then defined at points of the same parameter value $\mathbf{x}$. Thus, for two corresponding shapes, the following holds true for any parameter value $\mathbf{x}$:

$$\mathbf{S}_i(\mathbf{x}) \sim \mathbf{S}_j(\mathbf{x}), \tag{5.6}$$

where $\sim$ denotes correspondence.

Having established this notion of correspondence in parametric space, we can consider that for two unaligned shapes in parameter space, with one shape serving as a reference, the other shape can be brought into correspondence through *re-parameterisation*. We denote Φ_i as a re-parameterisation function for the i^{th} shape. Then, the re-parameterisation can be formulated as:

$$\mathbf{x} \overset{\mathcal{X}_i}{\longmapsto} \mathbf{x'} \doteq \Phi_i(\mathbf{x}), \tag{5.7}$$

where Φ_i is a diffeomorphism of the parameter space. The same mapping also affects the shape function $\mathbf{S}_i(\cdot)$, so that:

$$\mathbf{S}_i(\cdot) \overset{\Phi_i}{\longmapsto} \mathbf{S'}_i(\cdot),\ \mathbf{S'}_i(\mathbf{x'}) \equiv \mathbf{S'}_i\left(\Phi_i(\mathbf{x})\right) \doteq \mathbf{S}_i(\mathbf{x}). \tag{5.8}$$

It is important to note that since only the parameterisation function is changed, the original shape remains *unchanged* in object space. This concept is visualized in Fig. 5.3, where re-parameterisation changes the position of sample points on the hand outline, while the shape of the hand remains the same throughout different re-parameterisations. By fixing the parameterisation of a reference shape and changing parameterisation for a second shape, we can bring the object surfaces into correspondence.

Given the above explanations, it becomes clearer why it is more elegant to establish correspondence through re-parameterization instead of moving sample points on the shape surfaces in object space. The latter would require a diffeomorphic mapping for each individual shape onto itself, which changes with each adjustment of correspondence. Instead, we perform a diffeomorphic mapping from each shape into parameter space X only *once*, namely when the shape parameterization is initialized (see chapter 5.2.1.2). During the matching stage, when the surface is mapped to a reference, correspondence is established through an update of the parameterization, which itself is calculated by a diffeomorphic mapping of the *parameter space onto itself*. This mapping is equal for all shapes across the training set and also does not change for each correspondence update, since the underlying topological primitive and its parameter space (e.g. the unit sphere in cases of shapes with spherical topology) do not change throughout the optimization.

What remains is having a means of generating diffeomorphic mappings Φ_i in the spherical domain. In Davies et al. [48], different methods for spherical re-parameterisation are mentioned, e.g. piecewise-linear, recursive, localized and Kernel-based representations of re-parameterisations. However, in our case, since we are using a further parameterization using spherical harmonics and the SPHARM formulation, we can adapt the parameterization by rotating the complex-valued weights of the Fourier basis functions, i.e. rotating the expansion parameters. By using this rotational property of harmonic theory, it is not necessary to recalculate the SPHARM coefficients after rotating the parameterization [139]. Given a parametric SPHARM surface

$$v(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_l^m Y_l^m(\theta,\phi), \tag{5.9}$$

we can realize the required diffeomorphism $\Phi(\mathbf{x})$ as a transformation which rotates the parameter net on the surface by the Euler angles (α,β,γ). The new coefficients $\hat{c}_l^m(\alpha,\beta,\gamma)$ can be calculated as [139]:

$$\hat{c}_l^m(\alpha,\beta,\gamma) = \sum_{n=-l}^{l} D_{mn}^l(\alpha,\beta,\gamma) c_l^n, \tag{5.10}$$

where

$$D_{mn}^l = e^{-i\gamma n} d_{mn}^l(\beta) e^{-i\alpha m} \tag{5.11}$$

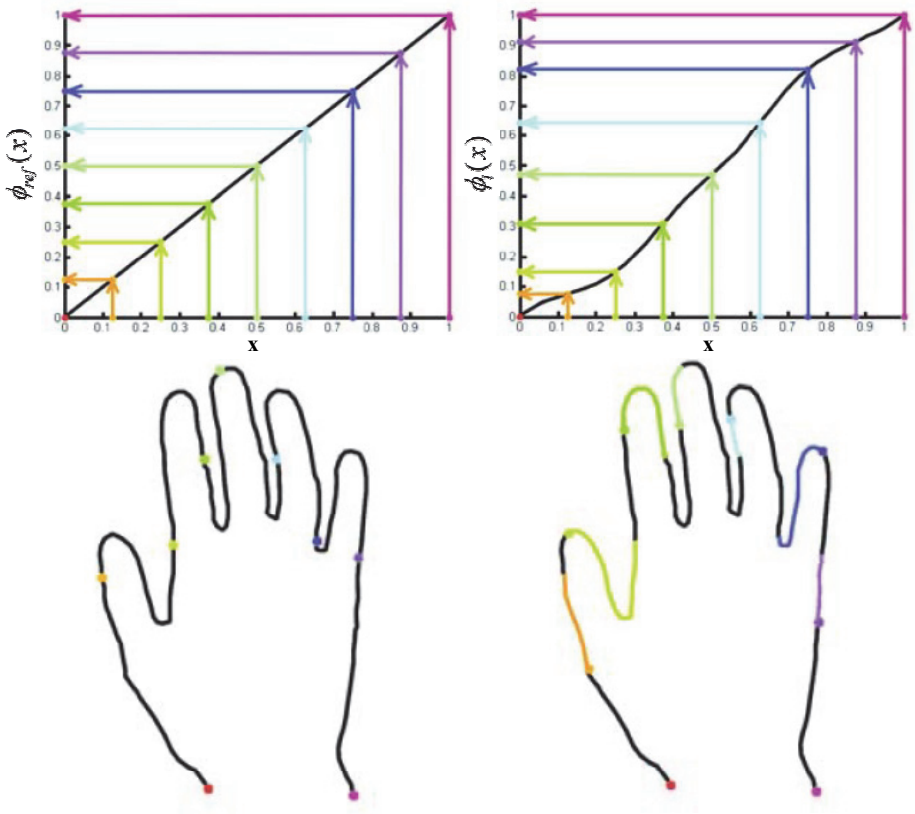

Figure 5.3: The concept of establishing correspondence through re-parameterization. The reference and moving shapes S_{ref} and S_i are represented by their respective parametric shape functions. The moving shape function $\mathbf{S}_i(\mathbf{x})$ can be changed through re-parametrisation using a diffeomorphic mapping $\Phi_i(\mathbf{x})$ in parameter space. In object space, this leads to a re-sampling of the shape surface at new locations, whereas the shape itself remains *unchanged*. Thus, instead of moving points on the shape surface directly, correspondence to the reference shape can be established through *re-parametrisation* (images adapted from Davies et al. in [48]).

and

$$d_{mn}^{l} = \sum_{t=\max(0,n-m)}^{\min(l+n,l-m)} (-1)^{t} \frac{\sqrt{(l+n)!(l-n)!(l+m)!(l-m)!}}{(l+n-t)!(l-m-t)!(t+m-n)!t!} \times \left(\frac{\cos\beta}{2}\right)^{2l+n-m-2t} \left(\frac{\sin\beta}{2}\right)^{2l+m-n}. \tag{5.12}$$

Objective function for model assessment: The objective function for comparing two SPHARM surfaces is calculated as the root-mean-square distance (RMSD) between the surface nodes or vertices. Due to the SPHARM representation, however, it is not necessary to transform the surface into object or Cartesian space. Instead, the RMSD can be calculated directly using the SPHARM expansion parameters of shapes S_i and S_j, i.e. their coefficients of the 3D spherical Fourier basis functions:

$$RMSD(S_i, S_j) = \sqrt{\frac{1}{4\pi} \sum_{l=0}^{L_{\max}} \sum_{m=-l}^{l} \left\| c_{i,t}^{m} - c_{j,t}^{m} \right\|^2}. \tag{5.13}$$

Spatial registration with ICP: Apart from re-parameterization, the *"SPHARM-MAT"* routine for establishing correspondence also performs an adapted form of the Iterative Closest Points (ICP) algorithm for adapting the spatial alignment (i.e. translation and rotation) with the reference. Together with re-parameterization, the overall registration of two shapes represented as SPHARM surfaces was coined "SHREC" by Shen et al. [141]. The additional ICP step is necessary to reduce linear mis-registrations as much as possible. In return, the final generated SSM discards these linear variations and only models the non-linear deformations of the target anatomy, which is the overall goal of SSM creation.

For translation and rotation of a SPHARM surface, it is not necessary to transform the shape back from parameter space into Cartesian space. Again, due to the SPHARM representation, these linear transformations can be performed in parameter space directly.

Given a translation matrix $\mathbf{T}$ and a rotation matrix $\mathbf{R}$, the complex-valued weights of the spherical Fourier basis function c_l^m can be re-calculated as [140]:

$$c_0^0(\mathbf{T}, \mathbf{R}) = c_0^0 + \mathbf{T} \times 2\sqrt{\pi}, \tag{5.14}$$

$$c_l^m(\mathbf{T}, \mathbf{R}) = \mathbf{R} \times c_l^m,\ l,\ m > 0. \tag{5.15}$$

Optimization: In the above paragraphs, we explained all necessary components for establishing pair-wise correspondence between an entire set of training shapes, given methods for registration in parameter space (re-parameterisation) and object space (ICP), as well as an objective function (RMSD). The only component left is an automatic optimization method for finding optimal correspondence according to the RMSD (eqn. 5.13).

In *"SPHARM-MAT"*, a sampling-based optimization strategy is pursued [139]: The reference parameterization remains fixed, while the parameterization of the moving shape is rotated to optimize surface correspondence by minimizing equation 5.13. The rotation space is sampled nearly uniformly using icosahedral subdivisions, creating rotation angles β and γ. If n denotes the number of icosahedral samples, the expansion coefficients $c_l^m(\alpha, \beta, \gamma)$ are rotated through $(0, \beta, \gamma)$ and then by n equal steps in α using equation 5.10. The RMSD is evaluated at each position in order to find optimal rotation parameters which minimize the RMSD. This procedure is iterated in a multi-level approach [141], where the best candidate rotations from one iteration are sub-sampled at finer angular resolution. Then, the optimization iterates until a minimal RMSD has been found.

5.2.1.5 Dimensionality Reduction

After performing the above described steps, a point-distribution model (PDM) is created, i.e. all training shapes have been sampled with the *same* number of vertices, where all vertices have been properly registered and are now *corresponding*.

At this stage, the M training shapes represent a distribution of shapes, where each shape is one sample in a $(K \cdot D)$-dimensional space. The last step of SSM building is to utilize a linear or non-linear dimensionality reduction technique to model and approximate this distribution.

The most common form of modeling the manifold of shapes is through linear approximation using Principal Component Analysis (PCA) [140, 68, 48, 43]. In this thesis, we also applied PCA for modeling the shape distribution.

The underlying assumption of PCA is that the distribution is Gaussian-distributed, i.e. elliptic in the high-dimensional shape vector space. While this is convenient in a mathematical way due to its linearity and also fast in terms of implementation and runtime, it can be well imagined that shapes are in fact not sampled from a Gaussian distribution. Instead, the distribution could be from a more complex, non-planar manifold in the space of all possible shapes. Therefore, non-linear distribution models have also been experimented with. A few will be mentioned in the following chapter 5.2.1.6.

The outcome of dimensionality reduction is a *mean shape*, along with M *"shape modes"*, which reflect the principal axes of deformation present in the training set. Furthermore, for each mode, shape variances can be calculated from the PCA eigenvalues as $\sqrt{\lambda_i}$. Therefore, the output of can be denoted as:

$$\mathcal{S}_\alpha = \mathcal{S}_\mu + \sum_{i=1}^{M} \alpha_i \mathcal{S}_i, \tag{5.16}$$

where $\mathcal{S}_\alpha$ represents a family of shapes which is parameterized by the mean shape $\mathcal{S}_\mu$, and a linear combination of shape modes $\mathcal{S}_i$, where each mode is weighted by mode-weights α_i. In chapter 5.3.1, we will describe how we use this parameterized family of shapes for midbrain segmentation in 3D-TCUS.

5.2.1.6 Alternative Approaches

The above sections described necessary steps for creating a SSM model according to the methods implemented in the comprehensive *"SPHARM-MAT"* toolbox. Naturally, there are alternative approaches and solutions for almost every step, with excellent overviews given in the review paper of Heimann et al. [68] as well as the book on SSM modeling by Davies et al. [48].

Alternative data representation: Instead of representing shapes as finite explicit models, e.g. triangular surfaces with vertices and faces, it is also possible to represent as infinite or continuous models. The underlying approach is to model shapes as zero-level-sets and to model shape variation e.g. as deformations of zero-level-sets to a reference zero-level-set. This approach is substantially different from explicit shape representation [68] and we refer the reader to the comprehensive review of Cremers et al. [43] for an introduction into level-set-based shape modeling.

Alternative surface parameterization: The choice of parameterization is mainly determined by the topology of the surface to be modeled.

Closed, genus-zero surfaces in 3D are the topological equivalent of the unit sphere and lend themselves to spherical paramaterization, albeit not necessarily with further parameterization using spherical harmonics. Gotsman et al. give an introduction into spherical surface parameterization methods [64]. Floater and Hormann [55] review several further methods for surface parameterization methods of spherical topology. A broader overview over mesh parameterization methods on the unit disk as well as the formation of more complex 3D base domains than the sphere is given by Scheffer, Praun and Rose in [137].

It is important to note that the above described SSM building method using spherical harmonics and the implementation in the *"SPHARM-MAT"* toolbox are restricted to closed surface of spherical topology. Open surfaces are the topological equivalent of a unit disk, and according parameterizations, and re-parameterization methods are required [55, 48].

Alternative methods for establishing correspondence: As pointed out in the review on explicit SSM creation methods by Heimann et al. [68], establishing correspondence is basically a registration problem. Accordingly, there are plenty alternative methods to solving the correspondence problem.

One approach is *mesh-to-mesh registration*, e.g. using one reference mesh (e.g. an average or atlas anatomy) and using linear and deformable mesh registration methods to transfer a moving mesh onto the reference. Subsequently, the reference vertices have to be mapped onto the moving surface, e.g. by piecewise-planar projection. Similarly, *mesh-to-volume registration* uses one reference mesh (e.g. from an atlas), which is then deformed into an unknown volume using e.g. gradient-snapping techniques, mostly to the binary segmented original volumes [68]. The prerequisite for this technique are robust and highly flexible deformable matching algorithm. Another alternative is

volume-to-volume registration, where e.g. an atlas-volume with pre-labeled landmarks is mapped onto an unknown volume using intensity-based deformable registration. Susbequently, the landmarks are transferred. As obvious from these examples, *mesh-to-mesh*, *mesh-to-volume* and *volume-to-volume* registration have the advantage that the correct number of landmarks is transferred and correspondence is established in the same pass. However, all three approaches depend strongly on the deformable registration algorithms, and in the case of intensity-based registrations (mesh-to-volume or volume-to-volume), on a target anatomy which has high contrast and clear distinguishment from the background. Also, it is not guaranteed that registered landmarks provide optimal correspondence. This, however, is the most important pre-requisite for the creation of an accurate shape model [68, 48].

A fundamentally different approach is to perform a group-wise or population-based optimization [48]. In this approach, all parameterizations are optimized at once, the cost function is often based on information-theoretic principles such as entropy [34, 33], minimum-description-length (MDL) [48, 47] or approximations of MDL [154]. The MDL principle is particularly attractive and achieves excellent correspondence and consequently SSMs of high accuracy [68, 47, 48]. It is based on the notion that the entire model is transferred as a "message" over a "channel". The objective function is based on minimum-description-length, which is a measure of the compactness of the model and thus the efficiency of the "channel transmission". The intuitive idea behind this approach is based on Occam's razor, which can summarized as the notion that "a simpler explanation is better than a more complex one" or more accurately, that among a selection of models, the ones making the "fewest assumptiones are usually preferable".

Alternative methods for dimensionality reduction: As mentioned, modeling the SSM with PCA assumes a Gaussian distribution of shapes, which can be expressed linearly by the combination of a mean shape with a weighted sum of orthogonal shape basis vectors. It is intuitive, however, that shape variations can be of significantly higher complexity than allowing for linear modeling using PCA. Also, the variation of PCA modes affect the shape in a global manner, changing all shape vertices at once. Other approaches are to use models covering more local variation, such as Independent Component Analysis [88, 68]. Also, several groups have investigated non-linear dimensionality reduction techniques and multi-variate distribution approximations, such as Gaussian mixture models, sparse PCA, maximum-autocorrelation factor, polynomial regression and Kernel PCA; for an overview, we refer the reader to [68].

Alternative for non-spherical topologies: The restriction to shapes of spherical topology is often prohibitive, e.g. for tubular structures, open surfaces and other non-closed anatomies. It is thus often desirable to have a more flexible framework.

Davies et al. briefly touch other topologies such as open surfaces (topologically equivalent to the unit disk) or 3D surface with holes (topologically

equivalent to a torus) and how to parameterize these objects in their book [48]. However, these parameterizations (and re-parameterizations) are also complex, in particular for the toroidal case and again not flexible if new topologies occur.

As mentioned, the approaches of *mesh-to-mesh*, *mesh-to-volume* and *volume-to-volume* registration can be theoretically of arbitrary topology. They are thus a good choice for quick and easy SSM creation experiments on arbitrarily-shaped objects. However, the template shape needs to be labeled carefully and deformable registration must be equally accurate for the resulting SSM model to be of sufficient quality.

In contrast to that, implicit shape representation using zero-level-sets can be inherently of arbitrary topology. This approach has a rigorous mathematical foundation, however it has to be noted that the signed distance maps representing level-set surfaces do not form a linear space, which can lead to invalid shapes if the variance in the training set is too high [68]. Again, we refer the reader to [43] for a proper introduction in such techniques.

An attractive hybrid approach is a *particle-system* based approach by Cates et al. [34], which is able to distribute discrete particles on a shape surface of arbitrary topology represented by a zero-level-set. This approach is elegant since it combines the flexibility of level-sets with the speed and tractability of discrete point-distribution-models. Furthermore, the objective function is entropy-based and thus along the lines of MDL, which achieves the most accurate SSM models up to now [68, 47, 48]. A starting point for experimentation with this technique is the Shapeworks toolbox implemented by Cates et al. and provided as open-source at [32].

5.2.2 Active Shape Model Segmentation

In the last section, we gave the theoretical background of building a statistical shape model (SSM). Our described workflow is one particular example of SSM generation, which is restricted to shapes with spherical topology and a parameterization using spherical harmonics on the mapping of shapes to their topological primitive, namely the unit sphere.

In this section, we want to give a short example and a basic overview of methods on how to apply SSMs and statistical appearance models (SAMs) for segmentation, turning them into Active Shape Models (ASM) [68] or Active Appearance Models (AAM) [61].

Due to the increased popularity of ASMs and AAMs over the past two decades, related literature on their application to medical image segmentation is vast and cannot be fully covered in this section. For a more complete overview on SSM-based segmentation, both on US and non-US data, we refer the interested reader to the comprehensive review by Heimann et al. [68]. This review alone lists 52 publications on segmentation methods using active shape models and active appearance models, as well as 22 other publications on non-segmentation applications of SSMs, including shape analysis, shape extrapolation and other applications like modeling breathing deformation or generating variable scenes for surgical simulation. The review by Gao et al. [61] provides further related literature, but solely dedicated to AAM model

building and their application for segmentation.

One of the earliest and most commonly applied approaches for ASM segmentation is to place an initial shape $\mathcal{S}_{\text{init}}$, represented by the average shape $\mathcal{S}_\mu$ and a vector of shape deformation weights α, inside a 2D or 3D image using a transformation T:

$$\mathcal{S}_{\text{init}} = T\left(\mathcal{S}_\mu + \Phi \cdot \alpha\right), \tag{5.17}$$

where $\Phi = (\phi_1 \cdots \phi_M)$ is the matrix of PCA eigenvectors, i.e. the orthogonal basis of shape deformation vectors. We denote the vertex locations of the initial shape $\mathcal{S}_{\text{init}}$ with the vector $\mathbf{y}$. Please note that the initial shape is firstly placed such that it exhibitis high overlap with the actual target object in the image.

Then, each control point (e.g. each vertex) along the shape boundary is evaluated according to an objective function. The objective function uses the image intensity (or patch intensity, patch texture or other features) at the control point location and measures the fit of the control point to its current location. Furthermore, it calculates fits at several other "test locations" in close vicinity of the control point, e.g. along the surface normal vector. Then the location with optimal fit is chosen as the new location for the control point, creating a set of displacement vectors $\mathbf{dy_p}$ for all shape vertices. Since this creates shape vertices at new, "arbitrary" locations, a shape is created which does not belong to the learned SSM shape manifold. Hence, as a next step, the shape has to be re-projected to SSM space, and the closest legal shape has to be found.

For reprojection, the SSM shape $\mathcal{S}^k$ has to be iteratively adapted to these new vertex locations $\mathbf{y} + \mathbf{dy_p}$, where k denotes the iteration number. This is first performed by rigidly matching $\mathcal{S}^k$ to the new vertex locations, e.g. using a Procrustes analysis [127] or ICP [26], yielding a new transformation T^{k+1}. Next, the new deformation weights α^{k+1} are determined using least-squares fitting in parameter space of the SSM model [68].

This approach was originally proposed by Cootes et al. [41] and since then has become one standard approach for ASM segmentation. Since today, several extensions and refinement to this basic search algorithm have been proposed. These are summarized in [68]. A first approach is to employ a coarse-to-fine multi-level approach, increasing speed and robustness of the search. Other improvements include constraining the shape adaptation per iteration, e.g. by weighting tangential vertex motion lower than normal motion. Significant improvements can be achieved by excluding outlier displacements per vertex, since PCA SSM models suffer from global shape deformation for each shape mode update. Hence, single vertex outliers can have profound impact on the overall shape.

In our work, in contrast to this search algorithm, we employ a localized region-based objective function for calculation of vertex displacements. Instead of calculating individual vertex displacements and mapping them back to the manifold of legal shapes as defined by the SSM through least-squares-fitting, we use a gradient-descent optimization and iteratively adapt the shape mode weights vector α_k. This has two advantages. First, we reduce the dimensionality of the search space to be optimized from 3 degrees-of-freedom per vertex ($3N$, in our case on the order of 2000) to only the number of shape

deformation modes (in our case on the order of 10-20). Second, we ensure that the midbrain surface stays within the space of legal shapes throughout the entire optimization, in contrast to regularly used optimization methods, where all vertices are first optimized independently and only subsequently back-projected into the shape manifold through least-squares mapping. Staying within the manifold of legal shapes throughout the entire optimization seems to lead to improved convergence speed and a better overall fit to the image data, as we discovered in early feasibility experiments for our method. The next section will describe the region-based cost function, optimization strategy and other components in more detail.

5.3 Materials and Methods

In the following, we will describe a method for reliable segmentation of the midbrain in the area around the SN. The proposed approach is based on three components: Firstly, the generation of a statistical shape model. Secondly, the combination of this shape model with an active surface framework. Thirdly, the active polyhedron framework of [151] to implement the discrete surface evolution, including a custom gradient-descent based optimization method.

5.3.1 Midbrain SSM Creation

For SSM creation, we utilize the comprehensive SPHARM-MAT toolbox of [138]. The toolbox was slightly modified to take a set of training meshes as input and generate the SSM model in one function. The only modified parameter was the SPHARM degree for expansion and it was set to 15 for all meshes, otherwise default parameters as downloaded with the toolbox were used. A 5-fold cross-validation was performed, i.e. the 22 meshes were split into five groups and five SSM models were generated from four groups (training set) to perform segmentation and evaluation on the fifth group (test set), leaving 16 training shapes per fold and thus 16 eigenvalues/eigenvectors per generated SSM. As mentioned, the output of the SPHARM toolbox trained on $M+1$ training shapes is a SSM with M modes of variation $\mathcal{S}_i$, $i = 1 \dots M$:

$$\mathcal{S}_\alpha = \mathcal{S}_\mu + \sum_{i=1}^{M} \alpha_i \mathcal{S}_i, \tag{5.18}$$

where $\mathcal{S}_\mu$ denotes the mean shape and $\alpha_1, \dots, \alpha_M \in \mathbb{R}$ are the shape deformation weights. The shape model $\mathcal{S}_\alpha$ can be considered as a parametrized family of shapes, where a specific configuration is completely determined by the shape vector $\alpha = (\alpha_1, \dots, \alpha_M)^T$.

The midbrain SSM is visualized in figure 5.4. The middle column depicts the mean shape $\mathcal{S}_\mu$, i.e. the average midbrain which was learned from the training set, while the left and right columns depict shape deviations $\mathcal{S}_i$ in distance of one and two standard deviations from the mean. In this respect, the first row illustrates deformations along the most significant mode of variation $\mathcal{S}_1$, while the second and third row depict deviations along $\mathcal{S}_2$ and $\mathcal{S}_2$, respectively.

5.3.2 Active Surface Formulation and Evolution

In order to evolve the shape model towards the desired configuration, we iteratively minimize an active surface energy of the form

$$E(\mathcal{S}) = \int_{int\mathcal{S}} f_i \, dx + \int_{ext\mathcal{S}} f_e \, dx, \tag{5.19}$$

where $\mathcal{S}$ denotes the surface, $int\mathcal{S}$ the region inside $\mathcal{S}$, and $ext\mathcal{S}$ the region outside $\mathcal{S}$. As our shape model provides enough regularity itself we employ no additional regularizer, such as the surface area.

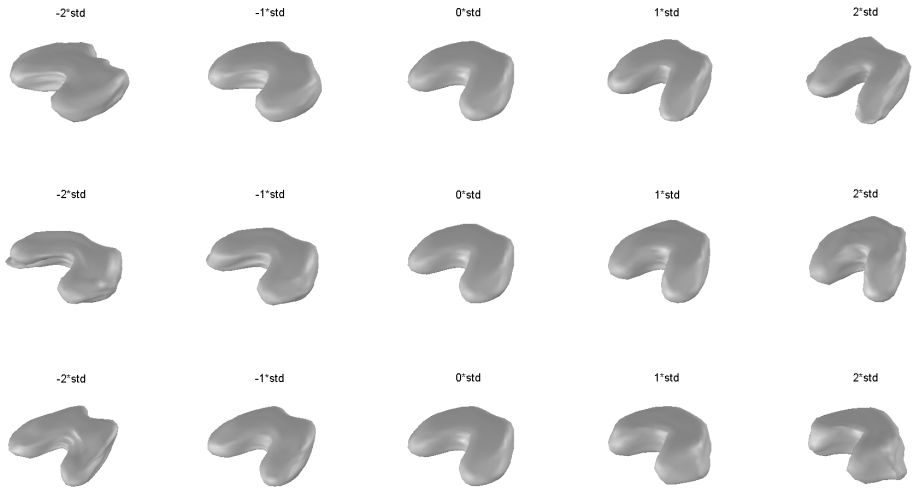

Figure 5.4: One of the five midbrain SSMs learned from the five cross-validation training sets. The middle column depicts the average midbrain, while left/right columns depict statistical modes of variation $\mathcal{S}_{\{1,2,3\}}$ (from top to bottom).

As the cost function, we use a region-based energy term defined by Chan-Vese [35]:

$$\begin{cases} f_i = (I - \overline{c_i}(x))^2, & \text{where } \overline{c_i}(x) = \frac{\int_{int\mathcal{S}} I(x)\,dx}{\int_{int\mathcal{S}} dx}, \\ f_e = (I - \overline{c_e}(x))^2, & \text{where } \overline{c_e}(x) = \frac{\int_{ext\mathcal{S}} I(x)\,dx}{\int_{ext\mathcal{S}} dx} \end{cases} \tag{5.20}$$

where $\overline{c_i}(x)$ and $\overline{c_e}(x)$ denote the average intensities inside and outside of the object boundary, computed globally over the image.

Due to the highly inhomogeneous nature of US images, however, foreground and background regions cannot be described by global statistics. For example, speckle patterns inside and outside the midbrain may occur in similar intensities and relative amounts across the whole image. Also, echo responses at anatomical boundaries often appear only locally, when tissue densities on both sides of the boundary are different from each other (see chapter 2.1.2 for more details on relevant US imaging physics).

Thus, we use a localized version of the Chan-Vese model proposed by [96]

$$\begin{cases} f_i = (I - c_i(x))^2, & \text{where } c_i(x) = \frac{\int_{int\mathcal{S}} B_\epsilon(x) I(x)\,dx}{\int_{int\mathcal{S}} B_\epsilon(x)\,dx}, \\ f_e = (I - c_e(x))^2, & \text{where } c_e(x) = \frac{\int_{ext\mathcal{S}} B_\epsilon(x) I(x)\,dx}{\int_{ext\mathcal{S}} B_\epsilon(x)\,dx}, \end{cases} \tag{5.21}$$

where $B_\epsilon(x)$ denotes a ball of radius ϵ centered at x. In oder to derive the evolution equation for the shape model, i.e. the gradient descent for the shape vector α, we plug the shape model (5.18) into (5.19):

$$E(\mathcal{S}_\alpha) = \int_{int\mathcal{S}_\alpha} f_i\,dx + \int_{ext\mathcal{S}_\alpha} f_e\,dx, \text{ with } \mathcal{S}_\alpha = \mathcal{S}_\mu + \sum_{i=1}^{M} \alpha_i \mathcal{S}_i, \tag{5.22}$$

and compute the partial derivatives with respect to the shape mode weights α_j:

$$\frac{\partial}{\partial \alpha_j} E(\mathcal{S}_\alpha) = \frac{\partial E}{\partial \mathcal{S}} \frac{\partial \mathcal{S}}{\partial \alpha_j} = \int_{\mathcal{S}} (f_i - f_e) N \cdot \mathcal{S}_j \, ds, \tag{5.23}$$

where N denotes the surface normal, and S_j denotes the vertex-wise cartesian shifts or deformations for the j-th shape eigenmode (i.e. the vertex "normals" of the j-th shape mode). Put together, these partial derivatives yield the gradient of E with respect to the shape vector α, which we denote by $\nabla_\alpha E$. In order to implement a gradient descent for α, we have to discretize the expression in (5.23).

5.3.3 Active Polyhedron Framework

We decided to use an explicit surface representation based on a triangular mesh as it is given by the *active polyhedron* method described by Slabaugh and Unal in [151]. Regular active surface models calculate the speed of vertex motion per iteration based on a cost-function which is derived from a local, regional or global statistic or descriptor. In [151], an active polyhedron is presented, which integrates the forces over the polyhedral faces and thus provides a lowpass filtering or smoothing over the surface as shape regularization. The regularization is provided by a cost-term which models forces between vertices similar to electro-static repulsion. Furthermore, the explicit framework is flexible enough to allow for complex mesh changes during the iterative optimization procedure, e.g. through edge splits, collapses, and face splits. This makes even topological changes of the shape possible, e.g. by introducing holes or splitting the overall shape into two shapes etc., leading to a quasi-implicit, level-set-like behaviour, while retaining the advantages of explicit formulation, such as improved speed or robustness.

However, as our created statistical shape model provides enough regularity we do not use the regularization described in [151] and evolve the model directly according to $\nabla_\alpha E$. This restricts the topological flexibility of the Active Polyhedron method, since the basic condition of spherical shape topology cannot be violated anymore. However, this restriction comes at the benefit of anatomical regularization and globally smooth surfaces without the need for domain-unrelated regularization models such as electro-static repulsion.

For the evaluation of the cost function, we need to determine which voxels are inside or outside a given mesh. This is trivial from an algorithmic perspective, however, it is the computational bottleneck of the optimization process as this task is performed for each iteration. We addressed this issue by integrating a simple but efficient GPU accelerated voxelization algorithm introduced by Crane et al. [42] into our framework. Once all voxels inside and outside the shape are determined, we can compute the local mean values c_i and c_e, which eventually allows us to approximate the expression in (5.23) as follows:

$$\int_{\mathcal{S}} (f_i - f_e) N \cdot \mathcal{S}_i \, ds \approx \sum_{k=1}^{N} \left[f_i - f_e\right]_k \left[N\right]_k \cdot \left[\mathcal{S}_j\right]_k, \tag{5.24}$$

where k denotes the vertex number and $[\cdot]_k$ denotes the evaluation at vertex k.

5.4 Experiments and Results

We performed the automated segmentation for all 22 subjects, given five folds of SSM models. The only user interaction necessary was the rigid placement of the mean shape $\mathcal{S}_\mu$ into the midbrain region. For all 22 experiments we kept the same parameter settings: the gradient descent step size was $\tau = 0.05$, the radius for the localization sphere $B_\epsilon(x)$ in (5.21) was set to 15 voxels ($\hat{=}$ 6.75mm), and the maximum iteration number was set to 100. Most experiments converged within this maximum iteration number, and typically required between 30-90 iterations until convergence. Convergence was reached when the RMS error between two consecutive shape vectors α_t and α_{t+1} dropped below 3% of the difference between the initialization shape vector α_0 (typically all zeros) and the first iteration α_1.

In order to demonstrate that the localization of the data term (cf. (5.21)) is necessary, we performed another series of experiments with a standard Chan-Vese model [35]. Further, we investigated whether preprocessing the image data with a few anisotropic diffusion steps yields better results, but we could not observe a significant improvement. In all quantitative evaluations, only slices with manual ground truth were considered, since those were the only slices in which the SN was clearly visible and manually segmentable by the medical expert, i.e. we restrict our evaluation to those slices with medical relevance to the diagnostic problem.

Regarding the midbrain segmentation, we calculated the Dice coefficient of overlap between ground truth segmentations and the segmentations obtained by our method. We can observe in Fig. 5.5 that using the localized data term improves the segmentation results significantly. Despite the poor image quality, the median of the Dice overlap of midbrain voxels across 22 subjects is 0.83, which means that in 50% of all cases the Dice coefficient is *at least* 0.83. In contrast to this, when the un-localized data term is used, we only achieve a median Dice of only 0.55, again supporting the assumption that global fore- and background statistics are not valid in the modality of ultrasound.

In order to evaluate the quality of the segmentation with respect to the diagnostic problem, we calculated the True Positive Rate (TPR) for SN segmentations, i.e. how many SN voxels were retained within the mesh after segmentation. We found that a median of 89% of SN voxels is retained by the midbrain ROI. We also post-processed all obtained segmentations with a dilation by one voxel ($\hat{=}$ 0.45mm), because we observed that the expert segmentation was systematically similar in shape to our segmentation, but segmented slightly further outwards, i.e. more into the hyper-echogenic regions surrounding the midbrain. In contrast, our segmentation converges slightly before these hyper-echogenic regions are reached. As we perform a ROI segmentation, a dilation of one voxel does not contradict the overall purpose. However, one should note that we only use dilation as one possible selection for a ROI, while our proposed method does not rely on dilation. This post-processing raises the median to 95%, i.e. half of our segmentations are able to retain more than 95% of voxels with diagnostic relevance, and five volumes achieve a perfect preservation of SN voxels after segmentation. Moreover, a dilation also

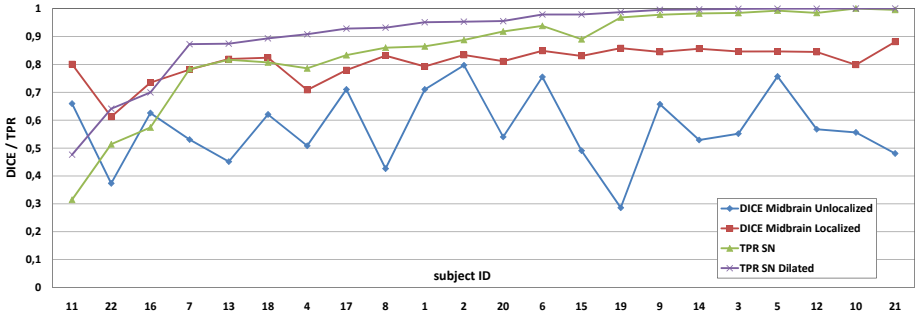

Figure 5.5: Dice coefficients for overlap of our automatic segmentation with the midbrain (left) and True Positive Rate (TPR) with the substantia nigra (right). Subject indices are ordered by TPR SN Dilated, the percentage of voxels with diagnostic value retained.

improves the automatic midbrain segmentation, yielding a median Dice value of 0.86, which again shows that our segmentation is very similar in shape to the expert opinion.

In addition to the quantitative evaluation presented in Fig. 5.5, we want to compare one of the poorly segmented cases (case 11) to two other cases with excellent (case 21) and medium (case 13) segmentation results (cf. Fig. 5.5), which serve as good examples for the performance of the algorithm on the remaining data sets. The upper, middle and bottom row in Fig. 5.6 show cases 21, 13, and 11, respectively. The last column of the figure shows mesh surface distance maps between the ground truth segmentation and the final automatic segmentation with localized data term, showing that the automatically segmented shape corresponds well with the ground truth. The reason for the relatively poor performance in case 11 can be explained by the comparatively bad image quality and the rather unusual shape of the midbrain, making also the manual segmentation very difficult - even for a medical expert. The same holds true for cases 16 and 22.

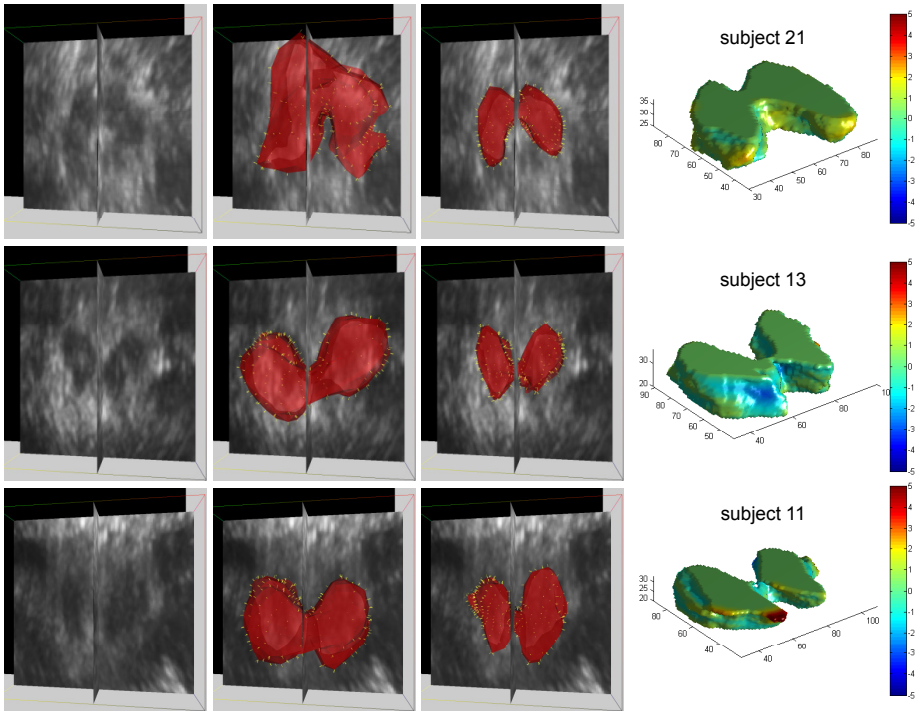

Figure 5.6: Exemplary Segmentation Results: Rows: data from three subjects. Columns: sample slice through volume with midbrain visible (left), segmentation result without data term localization (middle left), segmentation result with localization (middle right), mesh surface distance map between result and ground truth (colorbar in mm).

5.5 Discussion and Future Work

We have presented a robust and largely automated method for segmentation of the midbrain in 3D TC-US, which is, to the best of our knowledge, the first approach for *volumetric* segmentation of the midbrain from TC-US. The performed experiments clearly demonstrate that the segmentation performance is consistently high across 19 out of 22 subjects, although the quality of US volumes differs highly due to different thicknesses of the temporal bone windows, which proves the robustness of our method. As the image quality also depends on the used US settings and hardware, the localized region-based data term provides the advantage that is does not depend on these factors, in contrast to an Active Appearance Model for instance. Thus, the presented method can be readily applied to any form of 3D B-mode volume generation, such as wobbler probes or 2D matrix arrays.

In terms of usability, the proposed semi-automatic segmentation method reduces the overall segmentation time from approximately 20 minutes per patient for manual segmentation to around 1.5 minutes. As an accurate, robust, and user-friendly midbrain segmentation from 3D-TCUS is of high importance for a subsequent segmentation *and* classification of the SN, we believe that the proposed midbrain segmentation method is an important contribution to our overall proposed pipeline of computer-aided methods for PD diagnosis using TCUS. Most importantly, it lays the perfect groundwork for the segmentation of SN hyper-echogenicities, to which the following chapter will be dedicated.

6

Detection of Substantia Nigra Echogenicities in 3D Transcranial Ultrasound

6.1 Objective and Overall Approach

The past chapters have shown the value of transcranial ultrasound (TCUS) for diagnosis of neurological movement disorders and, in particular, for early detection of Parkinson's disease (PD). We have motivated the extension of the technique from 2D to 3D ultrasound, in order to improve upon often criticized shortcomings of the technique, such as objectivity and the high dependence on the sonographer's experience. We have shown first promising results towards computer-aided diagnosis using 3D-TCUS and manual segmentation of midbrain and substantia nigra (SN) hyper-echogenicities (SNE). Due to the impracticability of manual 3D segmentation, we have proposed a (semi-)automatic, computer-aided 3D midbrain segmentation technique in the last chapter.

In this chapter, we propose a method for fully automatic probabilistic detection and segmentation of SNEs in the midbrain. Since a manual segmentation of SNEs within the midbrain is similarly disruptive for the clinical workflow as manual midbrain segmentation, an automatic SNE segmentation marks another component of our proposed chain for computer-assisted PD diagnosis methods based on 3D-TCUS. As will be seen in the analysis of related work in chapter 6.2.2, a prior detection of the midbrain as a region-of-interest (ROI) is highly beneficial for the task of SNE detection and we are thus going to make use of the work from the previous chapter in the following.

The overall method for SNE detection is to perform a probabilistic classification of all voxels within the midbrain ROI using two random forest classifiers, one trained to detect hyper-echogenicities, the other inferring the probability of belonging to the SN based on its location within the midbrain. Only in combination, the two classifiers yield results close to human performance.

6.2 Related Work

Concerning related and previous work in literature, we will begin by giving a short overview of general lesion detection methods in 3DUS, in order to give a broad overview of work in the community so far. Subsequently, we will cite previous work on SNE detection in TCUS as a special form of lesions. Since there are no existing methods for quantitative SNE analysis in *3D-TCUS* so far, we will report on works by other groups on 2D-TCUS.

6.2.1 Lesion Detection Methods in 3DUS

There is a large body of related work for general detection of lesions in 3DUS. Applications include detection of cysts, lesions, tumors and other cancerous tissue types in various abdominal and pelvic organs, including breast, prostate, kidney, gallbladder, liver, and other anatomies [110]. Segmentation methods for tumor tissue are often designed towards large, contiguous, cyst-like tissue regions (with smooth or jaggy boundaries, but usually contiguous).

In contrast to that, SN hyper-echogenicities (SNE) in TCUS are often small and patch-like, since single speckle patches can already constitute a SNE. Due to this, most explicit active contour or active surface detection methods would fail, since previously defined contour topologies (e.g. as in snakes) would be too restricted to adapt to fine-scale and often disconnected speckle patches. Two approaches that are more applicable to our case are implicit active contours based on level-sets or methods for voxel-wise classification.

A plethora of methods enlisted in the reviews of Noble in 2006 [110] and 2010 [109] are based on these two general approaches, however apparently, no method has so far been applied to detect single speckle patches. On the contrary, speckle is often seen purely as noisy features which have to be removed prior to segmentation of larger anatomical structures. Therefore, many methods actually rely on speckle-reduction techniques such as speckle-reducing anisotropic diffusion (SRAD) [180] and other speckle filtering techniques [91, 92, 148, 153].

Furthermore, two difficulties we see with level-based approaches are that they would need some form of initialization within the midbrain at the location of the SN. A simple initialization from the midbrain outer surface and a convergence towards inner speckle patches would probably result in an over-segmentation and the inclusion of many false-positive patches within the midbrain, which anatomically cannot belong to the SN and thus should be discarded. The incorporation of such a spatial prior, however, is not straightforward in intensity-based level-set segmentation methods. The other problem is that the output of level-set segmentation is a binary map. However, given the difficult task of SNE detection, it would be preferable to obtain a fuzzy output indicating a voxel-wise confidence for the segmentation, leaving a final decision to the human observer.

Voxel-wise classification approaches often require no initialization, but classify lesions based on a combination of voxel-wise features, including intensity, intensity derivatives, local image phase or local texture [109]. The

methods stem mainly from the domain of machine learning and include methods such as linear and quadratic discriminant function, support vector machines, fuzzy inference system and Bayesian neural networks [109].

Similar to only some of these methods, our approach does yield a *probabilistic* map as output for SNE detection. A binary decision map can then be achieved by thresholding, however, the confidence level for each voxel is highly beneficial for SNE detection, especially since SNE segmentation can be challenging even for a human observer [161, 163, 162].

6.2.2 SN detection in 2D TCUS

There is little related work in literature so far concerning the automatic analysis of SNE in 2D, and no work at all on 3D-TCUS. However, similar to our work, all approaches we are aware of perform a midbrain ROI segmentation first and a SNE detection within the midbrain subsequently. Kier et al. [87] and Chen et al. [37] respectively perform SN pixel detection using morphological operators or image-feature-based SVM classification, both within a manually segmented midbrain in 2D. Engels et al. [52] use a hierarchical finite-element model and active contours to simultaneously segment the midbrain and SNEs in 2D. Sakalauskas et al. [132] analyze texture features within manually segmented midbrains in 2D for distinguishing PD from healthy controls. As mentioned, apart from our early contribution on midbrain segmentation in 3D-TCUS [2], there is no other previous work to our knowledge on (semi-) automatic midbrain or SNE analysis in 3DUS.

The main contributions of this chapter are therefore to 1) propose a novel and *volumetric* SNE detection method based on random-forests, 2) formulate a detection paradigm mimicking human experts by using probabilistic modeling of visual and spatial SNE features and 3) demonstrate the reliability of our SNE detection approach on our acquired 3D-TCUS dataset from 22 subjects (see chapter 3.4 for description of the dataset).

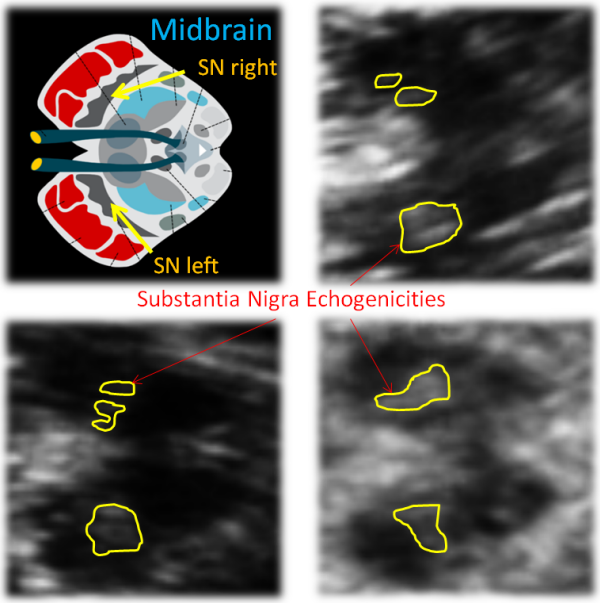

Figure 6.1: **Goal of our approach**: On the top left, the anatomy of the midbrain is detailed, showing the Substantia Nigra regions located at the front of both hemispheres. The other images show examples of typical SNE speckle patterns (in yellow) in 3D TCUS transversal slices.

6.3 Materials and Methods

As illustrated by Fig.6.1, an experienced observer can detect PD-related hyperechogenicities in the left and right SN using 3D TC-US. Unfortunately, TCUS cannot visualize the SN regions themselves, but only the high-contrast SNE speckles located randomly within the area of SN. Thus, relying on prior knowledge of the midbrain anatomy and the known rough location of the SN within the midbrain, an experimented observer has to decide whether an echogenicity belongs to the SN or not based on location and intensity of speckle patches. This makes the detection of Parkinson-related SNEs quite challenging. In this chapter, as mentioned, we aim at providing a reliable detection of PD-related SNEs in 3D by analogously integrating two types of information: (i) visual context and (ii) spatial location within the midbrain.

Since we use Random Forests (RF) as classifiers for voxels within the midbrain ROI, we begin by giving a short introduction into RF classification.

6.3.1 Random Forests Introduction

In the last few years, Random Forests have been shown to be state-of-the-art *ensemble learners*, which can be applied in many different tasks, including classification, regression, density estimation, manifold learning and semi-supervised learning [44]. The power of ensemble learning methods lies within two factors, namely (1) the combination of several weak learners into one

strong learner and (2) the injection of randomness during the model building stage, which significantly improves robustness and generalization of the overall model.

Specific advantages of random forests are: (1) being intuitive ("white box"), (2) fast and (3) highly scalable to large datasets. Profound coverage of this machine learning approach for classification and its recent generalization to above-mentioned tasks can be found in the technical report of Criminisi, Shotton and Konukoglu in [44]. While we refer the reader to that document for an in-depth coverage of the topic, we will reproduce a subset of the explanations from this report in the following. Alongside, we will provide a toy example, for better illustration of RF principles and preparation of the SNE detection approach explained in chapter 6.3.2. Basically, a Random Forest is an ensemble of decorrelated decision trees, which will be explained in the following.

Decision Trees: In [44], Criminisi et al. describe decision trees as "a set of questions organized in a hierarchical manner". The structure of a decision tree is an acyclic graph, with nodes and uni-directional edges, all in direction from the root node to the terminal nodes.

In a probabilistic way, the goal of a decision tree is to learn the posterior distribution that relates input observations to an output class. To this end, a decision tree uses a "divide and conquer" strategy. It first subdivides the observations into consistent subgroups, i.e. it creates a partition over the observation space, and second, it models the posterior locally in each part of the space. Finally, a piece-wise approximation of the posterior is obtained.

The subgroups are formed hierarchically by internal node or split node. Each node represents one question, or decision function. Depending on the answer to this question, the original data is pushed one level further down, where the next question is answered. In a classification tree, the outcome of the question is one option from a discrete set of classes $c \in \mathcal{C}$, with $\mathcal{C} = \{c_k\}$. This procedure of sequential questioning continues until the end of the tree is reached. The final nodes of a tree, analoguously to its nature-derived eponym, are called leafs. Each leaf stores a result in form of a class or a distribution of outcomes according to the training cases which arrived in that leaf during the learning process. In the testing stage, a previously unseen case, if passed down to a particular leaf, is classified based on the stored training cases stored in that leaf.

Formally, we denote an input data point as a multi-dimensional vector $\mathbf{v} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Each internal node is represented by a weak decision function with n-ary outcome. Assuming a binary outcome (as also assumed by Criminisi et al. [44] and by our approach in the following chapter), we can denote the following notation for the decision function at each node level :

$$h(\mathbf{v}, \theta_j) = [\phi(\mathbf{v}) \cdot \psi > \tau] \ \in \{\text{true,false}\}\,, \tag{6.1}$$

where θ_j denotes the decision parameters at node j.

As an example, we assume this decision function to be a single axis-aligned split in 2D, i.e. $\phi(\mathbf{v}) = (x_1 \ x_2 \ 1)^T$ and $\psi = (1 \ 0 \ \psi_3)$ or $\psi = (0 \ 1 \ \psi_3)$ for a horizontal or vertical split, respectively. This notation can be graphically

represented by a single-level decision tree (also often called a decision stump [159]). This thresholding decision, which is indicated by the bracket operator $[\cdot]$, separates the input space $\mathcal{S}$ into two output spaces $\mathcal{S}^L$ and $\mathcal{S}^R$.

Given an unknown data point, i.e. at the *testing stage*, the classification of the point using a decision stump is easy, since its class is determined by the binary decision defined in equation 6.1. For a decision tree with several levels, the *class posterior* is calculated instead:

$$p(c|\mathbf{v}) = \sum_j p(c|j)p(j|\mathbf{v}). \tag{6.2}$$

It remains to be defined how the decision tree is trained. At the *training stage*, a set of data points with known outcome or class, the so-called *training set* is presented to the tree. A set of decision functions θ_j is generated, either following a rule, or by *random* draws from a range of possible parameters. Each decision function θ_j is tested according to an *objective function*, which quantifies the quality with which the data is separated using this single decision. The best decision function maximizing the objective function *for this training set* is then selected and stored for the tree for future classification of unknown samples. As denoted by Criminisi et al. [44], most related literature employs the concepts of entropy and information gain as an objective function which chooses the optimal decision parameter from the available set. In our approach, we follow the same principle and we will give the definitions for entropy and information gain in chapter 6.3.3 (equations 6.8 and 6.7).

Having learned a decision model with "optimal" parameterization (at least "optimal" within the set of randomly drawn decision functions), an unknown data point can now be classified following equation 6.1. However, it is not hard to imagine that a single decision stump will often fail to yield correct classification, in particular if the data is not separable by an axis-aligned split. One approach to improve the learner is to allow a more "complex" decision functions, such as linear data separation using arbitrary lines in 2D or nonlinear data separation using conic sections [44]. Another approach, naturally, is to use more complex learners, e.g. by extending decision stumps to full decision trees with multiple levels. However, even then, a single classifier might not be sufficient for classification. Therefore, the concept of random forests is introduced.

Random Forests: A Random Forest is an ensemble of decorrelated decision trees. As mentioned, this concept is also called "ensemble learning" [27] and has been previously applied with large success, e.g. in boosting approaches [159]. The difference in RF ensembles is that a user-specified amount of randomness is introduced into the training stage for the weak learners.

The two most popular ways to introduce randomness are the following [44]:

1. *Randomized node optimization:* As mentioned, the parameters for the decision function can be sampled randomly from a pre-defined distribution, e.g. uniformly from a pre-defined range of threshold parameters. This

allows the training of trees on the entire available training data. Also, this approach achieves maximum-margin separation of the data, similar to Support-Vector-Machine models [45].

2. *Randomized training set sampling:* Instead of using the entire training set for training of each weak classifier, it is also possible to select a random subset of training samples. One possible approach is bagging, which yields greater training efficiency [44].

Using several weak classifiers trained with random node optimization or randomized sampling of the training set (or a combination thereof), the RF ensemble classifier can be used to jointly classify an unknown sample. Each of the T trees yields an output according to each tree's posterior distribution $p_t(c|\mathbf{v})$. The overall output can be calculated e.g. by averaging all individual tree posteriors:

$$p(c|\mathbf{v}) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|\mathbf{v}), \tag{6.3}$$

or by multiplying the tree outputs and normalizing with a factor Z:

$$p(c|\mathbf{v}) = \frac{1}{Z}\prod_{t=1}^{T} p_t(c|\mathbf{v}), \tag{6.4}$$

In summary, the most important parameters influencing a Random Forest classifier are the number of trees in the forest and the tree depth [44]. The tree depth is important for generalization, i.e. to make the tradeoff between under- and over-fitting. At the same time, the number of trees influences the smoothness of the posterior, and permits to increase the generalization further.

In figure 6.2, we give an illustration of a RF example based on decision stumps in order to visualize the most basic concepts of Random Forest ensemble classification. In figure 6.2, panel(A), we visualize a single decision-stump and one set of randomized decision parameters θ_A. In this example, the decision functions are line splits of the 2D plane. Furthermore, training samples are selected randomly from the whole available training set with a selection rate $\rho = 0.5$, i.e. only half of the training samples are selected randomly for each tree. In panels 6.2B and 6.2C, we visualize two additional, random binary stumps and their respective decision thresholds θ_B and θ_C. In panel 6.2D, we visualize an unknown sample in the 2D plane and how the individual decision tree (or stump) outputs are combined to form an ensemble opinion about the unknown samples' class c. In this toy example, the final random forest classification result is a fuzzy output, voting for class c_2, with a confidence of 0.67.

Our approach to SNE detection, as will be seen in the following section 6.3.2, is to model two properties of substantia-nigra hyper-echogenicities (SNE) as independent events, namely their intensity and location within the midbrain. Both properties are modeled with independent random decision forests, which follow exactly the properties and explanations above. The only differences are more complex decision functions, higher tree depth, and varying degrees of randomness, which will be explained in detail in the following.

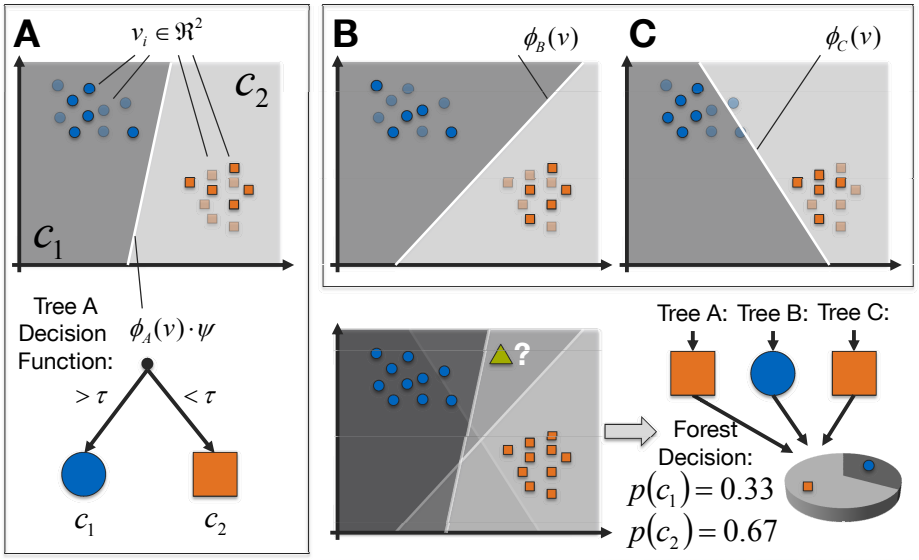

Figure 6.2: A) Visualization of single-tree learning with (random sub-selection of training set (selection rate $\rho = 0.5$) and general-oriented line splits as the decision function, B,C) multiple-tree learning with injected randomness and D) ensemble classification of an unknown sample using the RF classifier. As the weak learner, we selected simple decision stumps for explanatory purposes.

6.3.2 Problem Formulation

Let us consider an intensity function denoted by $\mathbf{I} : \Omega \rightarrow \mathbb{R}$, where $\Omega \subset \mathbb{R}^3$ is the image domain representing the 3D ultrasound data. We further assume that we are given a segmentation of the midbrain $\mathcal{M} \subset \Omega$, either from a manual expert segmentation or alternatively from the output of a ROI detection algorithm (see [2] and chapter 5.3).

In this paper, we propose to formulate the detection problem as a classification task in which each voxel $\mathbf{x} \in \mathcal{M}$ needs to be associated to a label $\mathbf{c} \in \{0, 1\}$, where 0 denotes the background and 1 the Substantia Nigra Echogenicities (SNE) class. In fact, $\mathbf{c}$ is the realization of 2 random variables $(\mathcal{E}, \mathcal{S})$ where $\mathcal{E}$ represents the observation of an echogenicity and $\mathcal{S}$ of the Substantia Nigra (SN), i.e. $\mathbf{c} = 1$ if and only if $\mathcal{E} = 1$ **and** $\mathcal{S} = 1$. Therefore, we aim at learning $P(\mathcal{E}, \mathcal{S}|\mathbf{x}, \mathbf{I})$, which represents the joint probability of observing an echogenicity $\mathcal{E}$ belonging to the SN $\mathcal{S}$ given the location $\mathbf{x}$ and the intensity function $\mathbf{I}$. It is important to note that:

1. it is not the SN itself which causes hyper-echogenicities but only potential acoustic micro-scatterers residing within it (see chapter 2.1.3) and

2. echogenicities can happen in the whole skull in TCUS due to tissue boundaries and micro-scatterers present in the entire brain tissue (see chapter 2.1.2).

Hence, we can assume the *independence* of the random variables $\mathcal{E}$ and $\mathcal{S}$,

and decompose this joint probability as follows:

$$P(\mathcal{E}, \mathcal{S}|\mathbf{x}, \mathbf{I}) = P(\mathcal{E}|\mathbf{x}, \mathbf{I})P(\mathcal{S}|\mathbf{x}) \tag{6.5}$$

The first term $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ is a data term, encoding the probability of observing an echogenicity given some visual information at location $\mathbf{x}$, and the second term $P(\mathcal{S}|\mathbf{x})$ is an anatomical prior not depending on $\mathbf{I}$, i.e. the ultrasound data.

As learning these probability distributions is challenging due to the dimensionality of the problem, we propose to use two discriminative models based on random forests. Following a "divide" and "conquer" strategy, random forests [30] provide efficient piecewise approximations of any distribution in high-dimensional spaces by: (1) partitioning the space using simple decisions, and (2) estimating the posterior in each "cell" of this space. As shown in [114, 44], random forests have been successfully applied to the task of multiple organ localization in CT and MR scans and further to organ segmentation in CT [107, 73]. Geremia et al. in [63] demonstrated state-of-the-art results for the segmentation of multiple-sclerosis lesions based on multi-channel MRI data.

In addition to a forest using *visual context*, we propose to learn a novel *spatial prior* based on two hemisphere-specific coordinate systems, which is thus perfectly adapted to asymmetric changes that can occur in 3D-TCUS concerning the scale and orientations of the midbrain structure. In the following, we describe how to use random forests for learning: (1) the data term $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ and (2), the spatial prior $P(\mathcal{S}|\mathbf{x})$.

6.3.3 Learning the Data Term

In TCUS, echogenicities are characterized by higher intensities and higher contrast. In this section, we describe how the data term $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ is learned . We propose to describe the visual context of a voxel at location $\mathbf{x}$ by extracting a set of simple features that encode the mean intensities in cuboidal regions of different sizes in the neighborhood of $\mathbf{x}$ similarly as described in detail in [63]. The regions are spanned by boxes of random location and random size in a large neighborhood of the target voxel. These regions can either be used as a local feature, where an array of cuboidal average intensities serves as a stand-alone feature vector, or as a context-rich feature, i.e. by sampling random box-pairs and using the differences of cuboidal average intensities as the feature vectors. As in [63], we use a combination of both for our experiments, namely 500 local features using random-sized cuboidal regions (edge length $[0.5 - 5]$mm) randomly placed in a comparatively large neighborhood around the voxel (range $[-1 \dots 1]$ cm) and additionally, 500 cube-pairs and their mean intensity difference as context features (same dimension ranges as above for the cubes).

Let us denote by $\mathcal{X}$ the space spanned by these simple features, and $\mathbf{X}$ the feature representation associated to a voxel at location $\mathbf{x}$. We consider a training set $(\mathbf{X}_n, \mathcal{E}_n)_{n=1}^N$, where each feature vector $\mathbf{X}_n$ is associated to a label $\mathcal{E}_n$ which is equal to 1 if there is an echogenicity at location $\mathbf{x}_n$ and 0 if not. Consisting of an ensemble of independent trees, a random forest permits to efficiently partition this high-dimensional space $\mathcal{X}$. Each tree can be seen as

a directed acyclic graph where each node consists in a decision function $f_{\mathbf{v},\tau}$ defined as:

$$f_{\mathbf{v},\tau}(\mathbf{X}) = (\mathbf{X} \cdot \mathbf{v} \geq \tau) \tag{6.6}$$

$\tau \in \mathbb{R}$ being a threshold and $\mathbf{v}$ being a vector of dimensionality $\dim(\mathcal{X})$ with only one non-zero entry, i.e. selecting a single axis or dimension from the many feature dimensions. According to the result of this decision function, incoming data are pushed towards the left or right child of the current node. Note that the role of $\mathbf{v}$ is to select a feature dimension where to perform the decision, yielding thus axis-aligned splits in $\mathcal{X}$. Please note that for each tree in the forest, only a percentage of the feature dimensions is used, in our experiments, we randomly select 10% or 100 dimensions for each tree, and perform axis-aligned splits at each node in the tree as described.

Let us denote by Δ the set of feature points from $\mathcal{X}$ reaching the current node, and Δ_l, Δ_r the subsets respectively sent to the left and right child nodes. As described in chapter 6.3.1, the choice of $\mathbf{v}$ and τ is optimized at each node following an objective function. In our setting, the optimization approach pursues a greedy strategy. A set Γ of functions are *randomly* drawn and the best candidate $(\mathbf{v}^*, \tau^*)$ is selected by maximizing information gain:

$$(\mathbf{v}^*, \tau^*) = \underset{(\mathbf{v},\tau)\in\Gamma}{\operatorname{argmax}}(\mathbf{H}(\Delta) - w_l\mathbf{H}(\Delta_l) - w_r\mathbf{H}(\Delta_r)) \tag{6.7}$$

where $w_l = |\Delta_l|/|\Delta|$ and $w_r = |\Delta_r|/|\Delta|$. $\mathbf{H}$ corresponds to the classical Shannon's entropy:

$$\mathbf{H} = -\sum_{e\in\{0,1\}} P(\mathcal{E} = e|\mathbf{x}, \mathbf{I}) \log\left(P(\mathcal{E} = e|\mathbf{x}, \mathbf{I})\right). \tag{6.8}$$

The posterior distribution can then be estimated from the set of points in the current node as:

$$P(\mathcal{E} = e|\mathbf{x}, \mathbf{I}) = \frac{|\{\mathbf{X}_n \in \Delta,\ \mathcal{E}_n = e\}|}{|\{\mathbf{X}_n \in \Delta\}|} \tag{6.9}$$

By optimizing this energy function, the tree aims at minimizing the uncertainty on the random variable $\mathcal{E}$, encouraging thereby the creation of leaves containing either mostly echogenicities, or mostly background. Nodes are grown until a maximal tree depth has been reached, or when the number of feature points falls below a given threshold. Finally, in each leaf, the posterior distribution $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ is computed on the set of features points reaching this leaf using Eq.6.9 and stored.

Now, to predict the probability of observing an echogenicity at a location $\mathbf{x}$ for an unseen ultrasound volume of the midbrain, one just needs to first extract its associated feature vector $\mathbf{X}$, to push it downward the tree until it reaches a leaf, and to use the stored posterior distribution. Considering a random forest consisting of T trees, ensemble model is to compute forest predictions by simply computing the average of all tree posteriors:

$$P(\mathcal{E}|\mathbf{x}, \mathbf{I}) = \frac{1}{T}\sum_{t}^{T} P_t(\mathcal{E}|\mathbf{x}, \mathbf{I}). \tag{6.10}$$

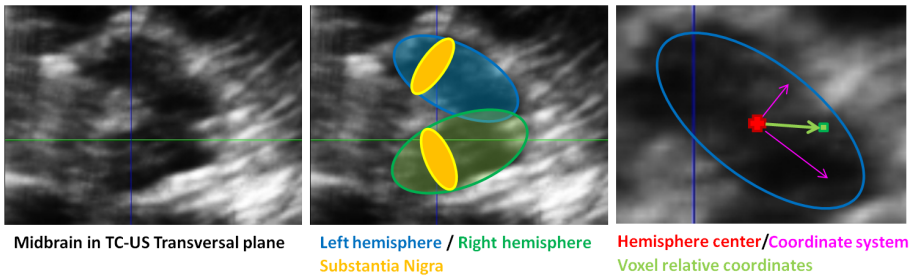

Figure 6.3: **Midbrain anatomy**: in the transversal plane, the midbrain has a characteristic butterfly shape. The Substantia Nigra are thin structures located at the front of both hemispheres. A hemisphere-specific coordinate system is computed in order to express and learn the spatial location of voxels within the midbrain, given their property of being a SN hyper-echogenicity or not. Due to the patient-specific encoding of spatial location, we can account for inter-patient asymmetric changes of scales and orientation.

6.3.4 Learning the Anatomic Spatial Prior

As shown on Fig.6.3, the midbrain has a characteristic butterfly shape in the transversal plane, which does not vary much along the longitudinal axis. The Substantia Nigra are thin structures located at the front of both hemispheres and do not vary much along the longitudinal axis either. Hence, we propose to express the location of each voxel using patient-specific coordinate systems that represent the left and right midbrain hemispheres in the transversal plane. By doing so, we can easily account for asymmetric changes of scales and orientation of the midbrain anatomy, which can occur in TCUS imaging.

Let us denote by $\{\mathbf{x}_m\}_{m=1}^{M} = \mathcal{M}$, the finite set of cartesian volume-coordinates for M voxels belonging to the midbrain. First, the coordinate centers of the left and right hemispheres are computed by performing a K-means clustering [27] on $\mathcal{M}$. Then, each voxel is associated to its nearest cluster center to create the 2 hemisphere subsets $\mathcal{H}^{\text{left}}$ and $\mathcal{H}^{\text{right}}$. Finally, principal component analysis on the set of voxel-coordinates is applied to each of these subsets in order to compute a hemisphere-specific transversal coordinate system, and the location of each point is expressed in the normalized coordinate systems of the hemisphere it belongs to. The in-plane location of each voxel $\mathbf{x}_m$ can then be encoded by a vector $\mathbf{x'}_m = [x'_m,\ y'_m,\ h_m]$, where x'_m and y'_m are the in-plane components in the hemisphere coordinate system, and h_m is a categorical variable encoding the left/right side.

To summarize, each voxel $\mathbf{x}_m$ is associated to a couple $(\mathbf{x'}_m,\ \mathcal{S}_m)$ for the training phase, where $\mathcal{S}_m$ is equal to 1 if $\mathbf{x}_m$ belongs to the Substantia Nigra and 0 if not. As in the previous section on visual prior calculation, we use a random forest to learn the spatial prior $P(\mathcal{S}|\mathbf{x})$ using a training set of 3D TCUS from different patients. During the training, each tree aims at separating the SN from the rest of the midbrain, and creates clusters in its leaves that are consistent in terms of spatial location $\mathbf{x'}_m$.

6.3.5 SNE detection

Once the data term and the prior have been learned from a set of labeled midbrains, a new unseen patient data can be processed as follows:

1. the midbrain is segmented,
2. the hemisphere coordinate systems are determined using K-means followed by a PCA,
3. the probability $P(\mathcal{E}|\mathbf{x}, \mathbf{I})$ and the prior $P(\mathcal{S}|\mathbf{x})$ are computed for each voxel, and
4. the joint probability $P(\mathcal{E}, \mathcal{S}|\mathbf{x}, \mathbf{I})$ can be predicted using equation 6.5.

Hence, we obtain for each voxel a probability of belonging to an SNE, and we can use a threshold $\mathcal{T} \in [0,\ 1]$ to create a binary segmentation of the ferrite deposits: $\mathbf{c} = 1$ if $P(\mathcal{E}, \mathcal{S}|\mathbf{x}, \mathbf{I}) \geq \mathcal{T}$, and $\mathbf{c} = 0$ otherwise.

The above methods were implemented in a custom toolbox, based on MATLAB and mex-functions (C++) for accelerated processing.

6.4 Experiments and Results

In this section, we evaluate our SNE detection approach on the bi-lateral 3D-TCUS dataset of 22 subjects described in chapter 3.4. We remind the reader that the 3D volumes were labeled by a blinded expert into the regions "midbrain", "SNE left" and "SNE right". For our validation, we will consider this manual labeling as gold standard.

We conduct comparative experiments to evaluate our SNE detection approach based on 2 discriminative models (VisForest-PriorForest) against the simple forest without spatial prior (VisForest), and a forest with a spatial prior constructed using a Gaussian model for each hemisphere (VisForest-GaussianPrior). We perform a leave-one-patient-out cross-validation, i.e. we train all models on 21 labeled midbrains and test on the remaining one. As the outputs from our system are probabilities between 0 and 1, we perform a ROC analysis, i.e. we vary the threshold's value to compute a binary segmentation, compute the corresponding confusion matrices for each run and derive different quality measures: f-measure, specificity and sensitivity.

The number of trees is set to 10 for all experiments, and best results were obtained for a depth = 15 for the VisForest, and for a depth = 10 for the PriorForest. Overall results are presented in Tab. 6.1. On the left, the best f-measure are reported by using threshold values of 0.5, 0.1 and 0.2 respectively for the VisForest, VisForest-GaussianPrior and VisForest-PriorForest models. By including our hemisphere-specific spatial prior, the f-measure is increased from **0.456** (VisForest) to **0.518** (VisForest-PriorForest).

Moreover, learning this prior distribution using a random forest provides slightly better results than with Gaussian prior achieving **0.508**. On the right, the best compromise between sensitivity and specificity are computed from the ROC analysis for all approaches. As illustrated by Fig. 6.4, the proposed prior permits to achieve improved specificity by better rejecting echogenicities that do not belong to the estimated SN. By varying the segmentation threshold, we also compute the area under curve which is **AUC = 0.903** for our approach, compared to a VisForest alone **AUC = 0.879** or with a simple Gaussian prior **AUC = 0.891**.

With a sensitivity and specificity of around **83%**, our SNE detection method is relatively close to the human inter-rater observability reported for experts segmenting in 2D (ICC 0.85) [99]. For this reason, despite the challenging nature of the data, these results are highly promising. Figure 6.5 shows a 3D visualization of one exemplary SNE detection result by our approach, also illustrating the promising performance.

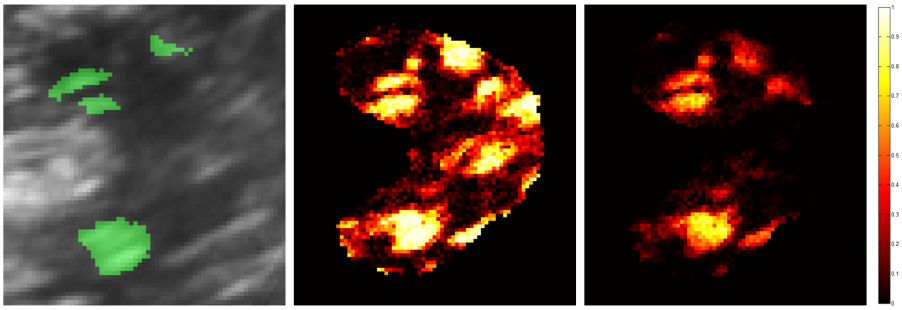

Figure 6.4: **The effect of our spatial prior**: From left to right, (i) the manual segmentation overlayed on the US data, (ii) the predicted posterior using the data term forest and (iii) the output after combining with the forest-based spatial prior. All outputs are probabilistic and can be thresholded to provide a binary segmentation.

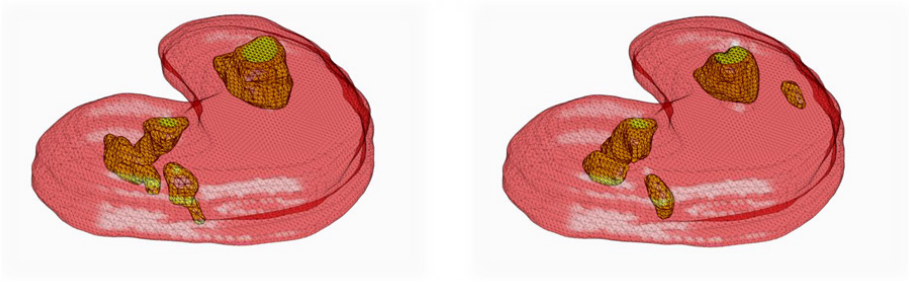

Figure 6.5: Results of Random Forest based detection of SNEs in 3D. Left: expert annotation. Right: our approach.

	F-measure			Specificity			Sensitivity		
	Mean	**Std**	**Median**	**Mean**	**Std**	**Median**	**Mean**	**Std**	**Median**
VisForest	0.456	0.115	0.463	0.775	0.060	0.779	**0.845**	**0.081**	**0.859**
VisForest-GaussianPrior	0.508	0.155	0.547	0.819	0.045	0.812	0.829	0.113	0.844
VisForest-PriorForest	**0.519**	**0.148**	**0.574**	**0.835**	**0.043**	**0.832**	0.828	0.099	0.829

Table 6.1: Overall SNE Detection results on 22 patients: The proposed prior permits to achieve better detectíon by improving the specificity, i.e. by better rejecting echogenicities that do not belong to the estimated SN. Moreover, using a forest-based prior provides slightly better results.

6.5 Discussion and Conclusion

In this chapter, as well as in our associated publication in [113], we presented the first approach for the automatic detection of Substantia Nigra Echogenicities in 3D TCUS. Since the interpretation of such data is very difficult and yields high inter and intra-observer variability (see chapter 4.3.4), our aim is to provide an objective and reliable segmentation of such Parkinson-related speckle patches.

Inspired by the way medical experts recognize SNE, we proposed a probabilistic formulation combining two discriminative models:

1. a ”visual” random forest specialized on the detection of echogenicities and

2. a ”spatial” random forest modeling a location prior within the midbrain.

For the spatial prior, voxel locations are parametrized within hemisphere-specific coordinate systems in order to account for asymmetric changes of orientation and scale in the midbrain anatomy. Through first experiments conducted on our clinical dataset with 22 subjects, we demonstrate the potential benefits of our approach in terms of segmentation accuracy, compared to expert labeling. From the segmentation output of our system, we can now quantify automatically the amount of hyper-echogenicities in each hemisphere. This marks another component in our overall chain of methods for computer-assisted PD detection using 3D-TCUS.

As a closing remark, we would like to point out that in this work, we made use of B-mode ultrasound data for SNE classification. However, it would be highly interesting to perform SNE analysis in the information-richer raw radio-frequency (RF) ultrasound modality (see chapter 2.3).

As summarized by Noble in [109], many lesion detection problems were solved using probabilistic approaches which were applied to the task of tissue classification in RF ultrasound. Although the number of such works is still limited, it is likely that it will be increasing in near future. This is due to the fact that in recent years, several ultrasound machines with research SDKs have become commercially available. RF ultrasound contains higher amounts of information about the back-scattered tissue echoes, since high-frequency components have not been filtered out yet, data re-scaling and 8-bit compression have not yet been applied and overall, the echo statistics have not yet been falsified by proprietary image enhancements often implemented in clinical ultrasound machines (see chapter 2.3 for details on the B-mode image processing chain).

At the end of this thesis, in chapter 8, we will describe our ongoing and future work concerning the recording of a new dataset, which will feature such RF data, and which we hope to analyze towards RF-based SNE detection.

7

Joint Segmentation-Registration-Reconstruction of Multi-View Transcranial 3DUS Data

7.1 Problem Statement and Objective

As described in chapter 3.1, we acquire 3DUS volumes of the midbrain in a bi-lateral faschion, i.e. by recording US data from the left and right bone windows, and combining them into a single volume during reconstruction. Another way of looking at this circumstance is that we are in fact dealing with an enhanced type of 3DUS acquisition, namely an extension from single-view 3DUS (SV-3DUS) to multi-view 3D US (MV-3DUS).

In related literature, MV-3DUS is defined as the combination of information from several SV-3DUS volumes into one reconstruction, which often bears several advantages. By combining several SV-3DUS volumes, the US-typical limited field of view can be extended, noise and speckle can be reduced and missing information from artifacts such as shadows can be reasonably compensated for. However, MV-3DUS introduces several challenges, as it is not rightaway obvious how to optimally combine US data from individual views.

In this thesis so far, and in our backward compounding reconstruction methods (see chapter 3.3), we have ignored these challenges. Therefore, the objective in this chapter is to utilize the image data from both bone windows and the fact that they have a large overlap of information, in order to create an improved reconstruction of the midbrain area. As we will explain later, we will use a sequence of inter-dependent segmentation-, registration- and reconstruction-steps to accomplish this.

7.2 Related Work in Multi-View 3DUS Reconstruction

One way the community has approached MV-3DUS in the past is through direct deformable registration of 3DUS using various distance measures [165], voxel-wise features computed from intensities [56], or variational approaches [183]. Another major approach is through joint registration-reconstruction. For example, [115] calculate a spatio-temporally smooth deformation field of different 3D+t echocardiographic views onto a common fusion space. In [178], several views of femur scans in fetal 3DUS are registered rigidly using image intensities and NCC as a distance measure. Furthermore, in [93, 122], a block matching approach is applied in order to improve the quality of several compounded images. To achieve this, Kruecker et al. [93] evaluated several similarity measures, concluding that sum of squared differences (SSD) is more suitable for low noise levels, whereas Poon and Rohling [122] focus on NCC for registration. Yu et al. [179] registered several view volumes of 3D echocardiography using segmented ventricles with the assumptions of rigid registration and no deformation.

It is common to all those approaches that the registration is purely intensity-based, which is prone to fail in presence of strongly varying speckle, and if SV-3DUS volumes differ significantly to each other, e.g. due to imaging artifacts such as shadows.

In this chapter, we are dealing with another fundamental problem of MV-3DUS. Depending on the type of 3D imaging system, e.g. 3D Freehand US or 2D matrix array, as well as depending on physical properties of the penetrated tissue, e.g. speed of sound, a combination of linear and non-linear spatial deformations can occur independently for each SV-3DUS. These varying distortions have to be properly compensated for, before a reconstruction with advanced methods as mentioned above can be performed.

The problem is formalized as the combination of several independent 3DUS volumes with individual spatial deformations into a single consistent reconstruction. This is achieved by simultaneously estimating and compensating the individual deformations through an innovative and inter-dependent sequence of segmentation-, deformable registration- and reconstruction-steps. Furthermore, we incorporate an anatomical prior into the reconstruction using the midbrain statistical shape model (SSM) (see chapter 5.3.1), which in turn guides the deformable registration.

We remind the reader that TCUS is generally of quite challenging nature due to the low SNR as a result of the US transmission through the skull bone with low frequency. This results in low SNR, large speckle patterns and a strong degradation of image quality with penetration depth, making intensity-based registration difficult. More importantly, scanning through the skull bone introduces non-linear shifts due to the irregular bone shape, different speed of sound in bone as well as non-linear diffraction at the bone-tissue interface, which we try to compensate using our method.

7.3 Joint Segmentation-Registration-Reconstruction Approach

7.3.1 Arrangement of JSR2 pipeline in a feedback loop

Our method for joint segmentation registration and reconstruction (JSR2) combines several single-view volumes of the same anatomic target object into an improved reconstruction using multi-view information.

We denote each SV-3DUS volume view as V_n^i, where the subscript $n = [1 \dots N]$ denotes the view index and N the number of views. The superscript i denotes the current reconstruction step iteration. Each volume can be considered to be residing within its own space $\Omega_n \subset \mathbb{R}^3$, where $\mathbb{R}^3$ is the real cartesian space in which our target anatomy is located. This means that for each V_n^i, there is a subspace Ω_n, created by the overall imaging system and thus affected by a groundtruth distortion or displacement field $\mathbf{U}_n^{true}$, which is unknown and varying w.r.t. $\mathbb{R}^3$. Factors for distortion can be linear, such as calibration errors, or non-linear, e.g. differences in speed of sound of scanned tissue.
In the first iteration, each input SV-3DUS volume V_n^i is represented by regular 3D sampling grid $X_n^0 \in \mathbb{R}^3$, which represents locations where the imaging system measures intensities $I(X_n^0)$. JSR2 tries to approximately recover the per-volume displacement fields $\mathbf{U}_n^{true}\left(X_n^0\right)$ by sequentially segmenting the anatomical target object in each SV-3DUS and subsequently registering segmentations onto a fused segmentation, yielding deformed volumes V_n^{i+1}. This iterative process is depicted in figure 7.1. When re-arranging the individual steps, the same process can be depicted as an iterative, closed feedback loop (see figure 7.2), which better illustrates that the JSR2 process is in fact optimizing a global energy, even though it is not yet unified in a single cost function. In the following, we denote the segmentation, registration and reconstruction operations and their results respectively as:

$$\mathbf{S}\left(V_n^i\right) = S_n^i, \quad \mathbf{Reg}\left(S_n^i, S_{joint}^i\right) = u_n^i, \quad \mathbf{Rec}\left(V_1^i, V_2^i \dots V_N^i\right) = V_{joint}^i, \tag{7.1}$$

where S_n^i denotes a segmentation result such as labelmaps or segmented surfaces, S_{joint}^i denotes a fused segmentation and u_n^i denotes a dense deformation field distorting the sampling grid, i.e. to $\mathbf{X}_n^{i+1} = u_n^i\left(\mathbf{X}_n^i\right) = \mathbf{X}_n^i + \Delta\mathbf{F}_n^i$. JSR2 reconstructs or approximates the distortion field $\mathbf{U}_n^{true}$, by merging individual image information into a joint volume and accumulating the required deformations for this matching, such that eventually:

$$\mathbf{U}_n^{true} \approx u_n^i \dots \circ u_n^2 \circ u_n^1\left(X_n^0\right) \tag{7.2}$$

The iterative optimization tries to find an optimal deformation field u_n^i, such that the energy or distance between segmentations S_n^i and a joint segmentation

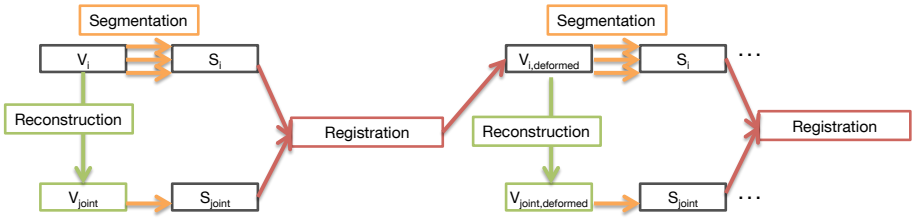

Figure 7.1: Iterative decomposition of JSR2 steps: Individual reconstructions from each view yield volumes in which the target anatomy is segmented. The segmented surfaces are used for feature-based, deformable registration. The deformation fields are applied to the original voxel positions where pixels from all 2D images are sampled and a new reconstructions for each view can be generated. Since these reconstructions have different voxel intensities than in the previous iteration, a new set of segmentations can be generated and the process can iterate.

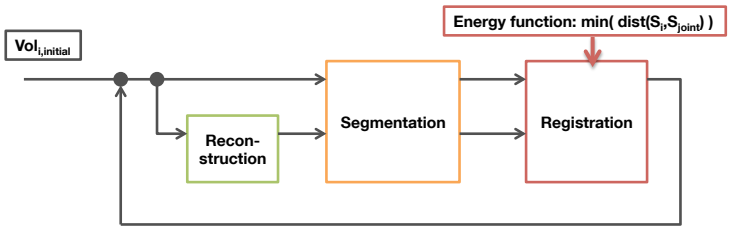

Figure 7.2: Visualization of JSR2 steps in a closed feedback loop: Compared to the iterative visualization, this visualization highlights that due to the closed loop, all steps affect each other and in effect optimize a global (although not joint) energy function.

is minimized, which happens in the registration step

$$\underset{u_n^i}{argmin} \left(\sum_{i=1}^{N} \left(dist \left(S_n^i, S_{joint}^i \right) \right) \right) \tag{7.3}$$

The effect of the procedure on the voxel sampling grids in the reconstruction step is illustrated in Fig. 7.3. Please note that under ideal circumstances, i.e. if segmentation and registration steps yield a perfectly congruent joint reconstruction, or if strongly simplified assumptions are made, e.g. only rigid displacements [179, 178], the iterative process will converge after a single iteration. However, if either segmentation or registration are too strongly regularized, or if the underlying displacements $\mathbf{U}_n^{true}$ are large, this process will iteratively drive individual volumes to a common reconstruction. In the following, we will give specifics for the algorithms $\mathbf{S}(\cdot)$, $\mathbf{Reg}(\cdot)$ and $\mathbf{Rec}(\cdot)$ used in this paper.

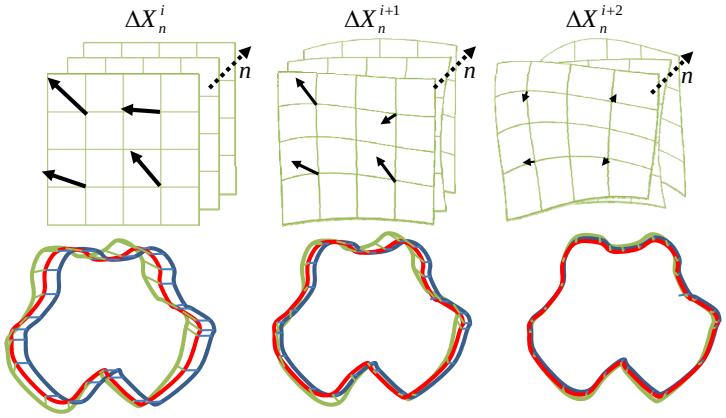

Figure 7.3: Illustration of JSR2 principle. Bottom row: The red shape represents the joint fusion of segmented shapes in individual SV-3DUS (green and blue). They are iteratively merged through deformable registration of segmentations, which yields new sample points $\mathbf{X}_n^{i+1}$ and a new reconstruction per iteration (top row).

7.3.2 Description of individual components

7.3.2.1 Segmentation method

For segmentation of the midbrain, we use our method proposed in chapter 5, which is based on a discrete active-surface model with shape prior regularization and a gradient-descent optimization of a region-based cost function with localized statistics. The 6-DOF rigid positioning of the average shape $\mathcal{S}_\mu$ in the volume is performed manually, at the beginning of the JSR2 procedure.

7.3.2.2 DeformableRegistration method

Registration can be performed by any deformable registration algorithm which is able to recover the deformation between the segmentations S_n^i and the joint segmentation S_{joint}^i. In this work, we decided to use a freely available method for dense deformable registration[1] based on a discrete Markov-Random Field objective function. We apply it to register voxelization of the meshes S_n^i, which we denote as labelmaps L_n^i. We use a multi-level registration approach, i.e. using a pyramid with three levels of control points and a control point distance of 1.4 cm on the coarsest grid, which partly allows for compensating small rigid components in the deformation field. The dimensionality of the optimization is reduced by using a set of control points and an additional discretization of the search space. At each control point, the registration cost between volume patches from the source and target volumes are calculated using any arbitrary unary distance function. Control point deformations are interpolated with a B-spline regularization in order to generate smooth and dense deformation vector fields $\mathbf{F}_n^i$. A smoothness term in the cost function additionally penalizes

[1] www.mrf-registration.net

large deformations of the grid through the MRF neighborhood relation of the control points. This allows the deformation from the segmentation boundaries to diffuse smoothly into neighboring regions. We use sum-of-absolute distance (SAD) on the labelmaps $L_n^i \sim S_n^i$ as the cost function to reduce the distances $D_n^i\left(S_n^i, S_{joint}^i\right)$. The registration method features a fast optimization approach based on efficient linear programming using the primal-dual principles **[?]**. After registration, the voxel sampling grids for each view are updated (i.e., deformed) using the deformation fields. The concatenation of deformation fields u_n^i in each iteration i are considered to eventually yield an approximation of the true deformation for each of the $n = [1 \dots N]$ US view separately, according to equation 7.2.

7.3.2.3 Reconstruction Approach

The original volumes in our dataset were reconstructed using a backward compounding technique [122]. We are working directly on reconstructed 3D B-mode volumes, assuming negligible reconstruction errors in the SV-3DUS volumes. We use the voxel grid in physical dimensions as our initial sample grid $\mathbf{X}_n^0$ for each SV-3DUS. Once SV-3DUS volumes are aligned to the consensus volume through deformable registration, we perform "reconstruction" by averaging all SV-3DUS volumes, i.e. $V_{joint}^i = 1/N \sum_{n=1}^{N} V_n^i$. This reconstruction is very simplistic, but serves for demonstrational purposes of our technique. Please note that since JSR2 is modular, more advanced reconstruction tecchniques can be applied here, for example in 3D Freehand US, ideally directly on the collected 2D pixel intensities in 3D space.

7.3.3 Simulation of 3D TCUS from MRI

In order to validate our method, we perform experiments on simulated midbrain 3DUS volumes from four different directions. In chapter 2.5, we described different approaches for US simulation, namely ray-based, wave-based or convolution-based. Due to sufficient realism and superior speed properties, we will make use of a convolution-based approach in the following.

As the name suggests, this type of simulation is based on the convolution of an artifical scatterer map with a sinc-shaped point spread function (PSF). The PSD simulates the system response of the US system in a simplified manner [8]. We simulate assuming 3MHz center frequency, a sinc-shaped PSF of 1 mm width in axial direction and Gaussian-shaped beam profiles in lateral and azimuthal directions (standard deviation 0.7 mm). As the artifical scatterer map, we take a cropped region of a midbrain in a T1-MRI volume and take image gradient components in x- and y-direction to simulate echo-producing tissue boundaries and different view directions. We affect gradient maps with uniform noise of 0.01% magnitude to simulate scatter variance in tissue, and distort the x- and y-map twice each with a random linear deformation (translation between $[-3...3]mm$, rotation between $[-3 \dots 3]^\circ$) and an additional non-linear random deformation on a $4x4x4$ grid (max. 2mm or 7% of the midbrain size). The simple simulation is very fast (0.9s for a

135x119x91 volume) and yields consistent speckle in all three spatial directions. The resulting midbrain volume is sufficiently realistic for the region-based segmentation algorithm to pick up midbrain boundaries.

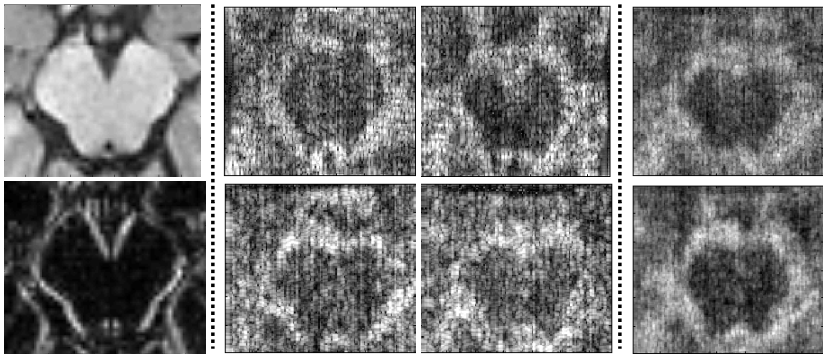

Figure 7.4: Results from simulated MV-3DUS. Left: midbrain crop from T1-MRI (top) and image gradient in x-direction as basis for scatterer map (bottom). Middle: four linearly and non-linearly distorted scatterer maps and convolution-based 3DUS simulation results (only single slices shown). Right: regular average compounding (top) and JSR2 reconstruction (bottom).

7.4 Experiments and Results

Validation using artificial data. Results from the simulated midbrain volumes from MRI can be seen in Fig. 7.4. The reconstruction with regular averaging, which is comparable to normal backward compounding can be seen in the top right. Although the midbrain shape can be seen, boundaries are blurred and image contrast is reduced. In contrast, the JSR2 reconstruction on the bottom right has sharper midbrain boundaries, better contrast, better preservation of symmetry and of features such as the ring of hyperechogenic aquaeducts surrounding the midbrain. For quantitative evaluation, we performed five random trials of this experiment and compare results to a simulation on the gradient map *without* deformation. Compared to a regular reconstruction by averaging, JSR2 showed a better NCC value (0.52 vs. 0.48, i.e. 8% improvement) and improved the estimated SNR (3.47 vs. 3.3 or 5% improvement). Furthermore, since we know the groundtruth deformation of our four SV-3DUS volumes in each iteration and each trial, we calculate the percentual deformation compensation on the segmented mesh boundary for all views in the final iteration of all trials. On average, the magnitude is reduced by 14.2%, demonstrating that the JSR2 method was able to partly recover the deformation maps.

Validation using clinical data We also compared the JSR2 reconstruction to a regular average reconstruction using in-vivo data on ten subjects, provided in the database. Naturally, there exists no gold standard for the mibrain deformation due to transcranial scanning, to which we can compare our reconstruction result. Instead, we present two reconstruction results using regular average compounding and JSR2 for qualitative comparison in Fig. 7.5. In the left example pair, the white arrow highlights that hyper-echogenicities in the Substantia Nigra (SN) region were reconstructed much more consistently,

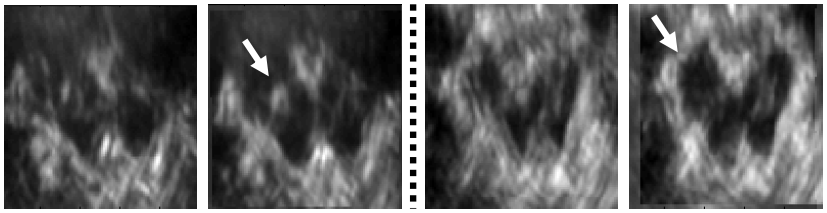

Figure 7.5: JSR2 results on real data. Left and middle left: bi-lateral reconstruction using regular average compounding (left) compared to JSR2 reconstruction (middle left). The right figure half shows another example pair.

which is an important feature for diagnosis of Parkinson's disease [168] (see chapter 1.3). In the right pair, the midbrain boundary is much more consistent and displacements in the average compounding are almost fully compensated. Overall, in five out of ten cases, reconstruction was improved similarly well as in Fig. 7.5. In two cases, the JSR2 result was slightly better than regular backward-compounding and in three cases, the reconstruction result showed no significant improvement. All JSR2 runs converged, the median number of iterations until JSR2 convergence was four (min: 1, max: 13).

7.5 Discussion and Future Work

The results section has shown that in simulated MV-3DUS with $N = 4$, JSR2 yielded reconstructions that are more symmetric, have sharper boundaries, higher contrast and better preservation of feature on or close to the midbrain boundary. Also locally, i.e. close to the midbrain, deformation fields could be partly compensated. In in-vivo transcranial MV-3DUS with $N = 2$, JSR2 yielded results that can qualitatively be assessed at least as good as average compounding or better, in some cases with significant improvement.

Naturally, JSR2 can only compensate deformations in close vicinity to the segmented boundary and thus the deformation away from the boundary quickly fades away. There are at least two ways to counteract: first, by using a combined registration method incorporating shape information close to the boundary and intensity-based registration further away, and second, by including more anatomical structures into the JSR2 process, e.g. by using coupled SSMs and multiple organs, which would offer larger spatial support for correct deformation of surrounding tissue.

Concerning the segmentation, we noticed that segmented boundaries were not always matching midbrains in SV-3DUS very well. The main reason for failure in SV-3DUS probably is that the provided SSM was trained on expert segmentations in bi-laterally reconstructed volumes, i.e. volumes which already contained reconstruction errors and thus do not represent midbrain anatomy in a fully correct way. We are convinced that if the segmentation worked more reliably on single-view 3DUS volumes, the JSR2 reconstruction results could have been even better. Unfortunately, since the individual JSR2 components are currently not regularised with a common penalising cost, a failed segmentation can have a large impact on the overall outcome, since falsely segmented midbrain surfaces are registered onto each other, leading to an overall distorted joint reconstruction. Also, if one of the SV-3DUS segmentation fails locally, and the mis-segmentation is close to the region of the SN, the diagnostically relevant part of the image would undergo large deformations. Since the currently best diagnostic features are based on volume measurements of SN hyper-echogenicity, this would have a large impact on the diagnostic outcome of the classifier. Again, a joint cost function for the three steps segmentation, registration and reconstruction, could lead to a better joint reconstruction, in particular since the overall deformation of SV-3DUS volumes could be regularised. It is probably also advisable to incorporate an additional data term in the cost function which considers image-based registration, e.g. using a cost-function such as normalised cross-correlation.

Another important shortcoming of our work so far is the partly qualitative evaluation of the method on clinical data, as described in the previous section 7.4. One important problem here is, as mentioned, that no groundtruth exists for the clinical data, as it is not known what kind of linear and non-linear deformations are actually present and what their magnitude is. For proper evaluation, we would require another image modality which needs to be registered to 3DUS for comparison. In the following chapter, we will describe our upcoming and future work, one part of which will be the acquisition of

a multi-modal 3D-TCUS dataset which includes co-registered MRI volumes (see chapter 8.2.2). This data can serve as a "gold standard" and allow us to perform more substantial, quantitative evaluation of the JSR2 approach in future.

In summary, the JSR2 approach proposed in this chapter is still facing several challenges and thus not yet readily applicable to the reconstruction of transcranial MV-3DUS volumes. However, based on the experienced gathered in the above-mentioned experiments, we are now able to better assess the challenges of this reconstruction problem and give indications to improvements, such as a joint cost-function plus regularisation and an additional image-based cost term. Insofar, the work proposed in this chapter is an important step towards an advanced reconstruction of this challenging data, and a good basis for future work.

8 Discussion and Future Work

In the past chapters, we have explained our efforts in extending the 2D-TCUS method for Parkinson's disease (PD) diagnosis to 3D, for increased objectivity and first *volumetric* analysis of SN hyper-echogenicities (SNE). We have described our system setup and data acquisition (chapter 3) and subsequently demonstrated first results towards computer-aided diagnosis using multiple observers and manual segmentation (chapter 4). For improved usage of the 3D-TCUS technique in medical routine, we have proposed two unprecendented techniques for computer-aided 3D segmentation of the midbrain (chapter 5) and of SNEs within the midbrain ROI (chapter 6). We have also proposed a first step towards improved 3D reconstruction of the midbrain area using a novel joint segmentation-registration-reconstruction approach (chapter 7). In the following, we will first discuss these results from the perspective of our original goals and contributions. Then, we will describe our currently ongoing work in this direction and what future steps we envision and recommend in order to further develop this research area and possibly bring it to clinical routine in future.

8.1 Discussion of Overall Results

The methods we proposed and the results we achieved and described in the last chapters reflect the work of a more than three year long inter-disciplinary collaboration between our group on the technical side, and our medical partners at Klinikum Grosshadern, Munich. Retrospectively, there are several thoughts and conclusion that we can draw from our achievements of the past years.

Availability of 3DUS acquisition in clinics: Our most important premise in this work is the *availability of a 3D ultrasound system* at the clinical site performing 3D-TCUS. A natural critique at our approach is that most clinics do not actually own a 3D transcranial ultrasound system. Currently, 3D systems based on 2D matrix transducer arrays (see section 2.4) are still too expensive and probably offer too little a resolution for quantitative analysis

of SNEs. Wobbler-based 3DUS systems are more widely available, but mostly designed for abdominal and obstetric applications, and thus mostly unusable for transcranial scans due to the large and curved transducer footprint (see section 2.4).

3D Freehand ultrasound systems, on the other hand, have superior image resolution, but are rarely available outside research collaborations. Commercial system include the SonoWand system [156] (SONOWAND AS, Trondheim, Norway), which was not originally marketed for transcranial examination and also seems to have ceased distribution by today. Another system is the CureFab One system (CureFab GmbH, Munich, Germany), which is based on electro-magnetic tracking. It is designed as an add-on device on arbitrary 2D ultrasound systems, which are mostly present at the clinic already. The low price of such an add-on device compared to a full 3DUS system might make this an attractive alternative.

In general, while we acknowledge the critique that systems for 3D-TCUS acquisition are not yet widely available in clinics, the general projection in research and industry is that such systems will become more and more commonplace in future, especially once 2D matrix arrays have improved in resolution and decreased in price. As mentioned in section 2.4, we would like to remind the reader at this point that all methods proposed in this thesis are also directly applicable to 3DUS acquired with 2D matrix arrays.

System setup and data acquisition: We have proposed a *3D Freehand ultrasound acquisition* system in chapter 3, with which we recorded 22 subjects, 11 healthy controls and 11 diagnosed PD patients. If the 3D Freehand ultrasound system is well calibrated and the reference target at the subject's forehead does not move, the obtained 3DUS volumes are of high quality, even if reconstructed bi-laterally, i.e. by combining image information from the left and right bone windows into one volume. With good calibration, the two volumes from both sides have actually surprisingly good overlap (see figure 3.5). Sometimes, this overlap is better than one would expect given certain physical effects, such as spatial mis-registration due to speed-of-sound differences between cranial bone and brain tissue.

However, in some cases, the bi-lateral volumes observe noticeable mis-registration (see figure 7.5). In our experiments, we could not always pinpoint the exact reason for these mis-registrations, in particular since the calibration parameters did not change in-between recording sessions. Possible reasons are (1) unwanted and undetected movement of the forehead reference target (e.g. due to a sweaty forehead, undesired mimic changes of the proband which moved the target or slight bumping of the examiner against the target during acquisition), (2) partial blocking of line-of-sight to a subset of the tracking markers, reducing optical tracking accuracy, (3) variation of physical ultrasound propagation across different patients (e.g. bone tissue causes larger mis-registration in some subjects than in others. Please note that this error would be very difficult to predict and quantify, especially without an additional imaging modality which highlights bone such as CT).

In general, while our recording setup yielded good results almost through-

out, it took us several iterations of *refining our recording setup* until stable results were obtained. Naturally, the setup is still not optimal at the current stage. A better, i.e. *more rigid fixation of a head reference target* would be desirable, however it should be still comfortable for the patient, and more importantly, the pre-auricular bone windows should not be obstructed by the head fixation. Our solution, a headlamp with attached optical tracking target, which is tightly fixated to the forehead, is a good first step and sufficient for research, but can certainly be improved for clinical routine.

Another improvement could be an *improved ultrasound calibration* routine. We use single-wall calibration (chapter 3.2), which we mainly chose due to its satisfying accuracy while the phantom itself is very easy and cheap to construct. Our setup did not require re-calibration for each acquisition. However, it would be very desirable if it was possible to perform a *calibration verification prior to acquisition*. This was not possible for our setup, since the single-wall phantom would move in-between scans, and verification with such a phantom takes too long for clinical routine. Other phantoms, such as z-phantoms allow for more rapid and accurate calibration and pre-scan verification. In appendix B.4, we describe the work of a Master thesis supervised during this PhD thesis, in which a z-phantom was designed and implemented. However, the phantom was not precision-manufactured and could not be applied in our experiments yet. The optimal solution, highly desired by us during our work, would be a commercial or high-quality open-source solution for 3D Freehand US calibration, with high accuracy, reliability, repeatability and speed. Unfortunately, commercial systems are not available, and open-source solutions are often out-dated or restricted to certain ultrasound machines, or research SDK versions (e.g. the StradWin system [62] does not support most types of framegrabbers, or the most up-to-date versions of the Ulterius research SDK by Ultrasonix).

Computer-aided diagnosis: The first results from our multi-observer study in chapter 4 and in our journal publication in "Ultrasound Med & Biol" [121] seem highly promising and encourage us to continue our efforts in the area of 3D-TCUS for PD diagnosis. Apart from the journal publication, our efforts were also awarded by the German Parkinson's Disease association [119].

However, it is important to note that our results were based on *manual midbrain and SNE segmentations*. In near future, we would like to combine our pipeline of developed methods and compare *diagnostic results based on fully computer-aided segmentation* with manual segmentations. Given the high regional overlaps and volumetric correlations consistently above 80%, when comparing our proposed segmentation methods with human expert opinions (see sections 5.4 and 6.4), we expect fully computer-aided diagnostic results to be also promising, and possibly more objective than given human observer opinions. We are currently obtaining *a third human-rater segmentation* for our 22-subject dataset, with which we plan to perform this analysis on a larger scale (see chapter 8.2.1).

Midbrain segmentation: The midbrain segmentation in our method achieves promising results, with Dice overlaps of 0.86 and 95% retained SN voxels

within the midbrain ROI. In our 22-subject dataset, it proved to be accurate, while being robust towards missing boundaries, large variation in overall contrast and anatomic detail due to varying bone windows. We also tried the midbrain segmentation on data collected with another ultrasound device (Sonix MDP, Ultrasonix, British Columbia, Canada) with similar qualitative success, although we have not yet performed a quantitative analysis yet.

A shortcoming is that our method depends on a statistical shape model, which we created from segmentations on bi-lateral reconstructions. Naturally, the model has picked up the particular appearance of the midbrain in *bi-lateral volume reconstructions*. Since bi-lateral reconstructions are inherently flawed due to linear and non-linear acquisition and reconstruction errors (see chapter 7.1), manual segmentation and consequently the SSM are flawed as well. For example, we have observed that segmentations in uni-lateral midbrain volumes (e.g. just from the left or right bone windows) using our method (see JSR2 segmentations, chapter 7.3) leads to noticeable mis-segmentations. In future work, we plan to build a *SSM from anatomically "correct" images*, e.g. from MRI.

Another shortcoming of the midbrain segmentation method is the initialization, which we performed manually in our experiments so far. An *automatic initialization* would be highly preferrable. In future work, we plan to experiment with Random Forest regression, similar to [114, 63], in order to predict the transformation matrix T_{init} for initial placement of the SSM in the image volume.

SNE segmentation: The Random Forest classification approach for probabilistic SNE detection yielded very good results, given the difficulty of the task. In this work, we used multi-level cuboidal averages as features for RF training and testing. Naturally, this approach can be expanded with more visual features. We have already started experiments using additional visual features, such as histogram of oriented gradients (HOG) [46], Laplacian-of-Gaussians (LoG) [66], local binary patterns (LBP) [114] or Gabor features [39]. Since random forests are inherently feature-selective due to a random selection of feature dimensions during the training stage, they can deal well with such an abundance of features. Initial experiments so far have shown that using a larger variety of features actually improves classification so much that it slightly *exceeds* the SNE detection performance reported in chapter 6 and our MICCAI publication [113], even *without a spatial prior*. Using an additional spatial prior, we expect much better performance than reported in this thesis. Moreover, we plan to conduct experiments towards *joint midbrain and SNE detection and segmentation* using a RF approach, possibly enabling us to skip our proposed SSM-based active surface method overall.

Improved 3DUS reconstruction of the midbrain region: In chapters 2.1 and 2.2, we have explained several physical properties of ultrasound wave propagation and tissue interaction which lead to linear and non-linear registration errors between the 3DUS scans from left and right bone windows.

In chapter 7, we proposed a joint segmentation-registration-reconstruction (JSR2) approach to remedy this problem. Our initial results using the JSR2

algorithm look promising, and we are motivated to further expand on this initial approach. However, our evaluation until now is very limited and we have observed certain instability of the feedback loop given unappropriate parameterizations of the segmentation and registration steps. Most importantly, the individual steps are still only performed sequentially. A *joint energy term* would be preferable, in particular in combination with a *joint regularization approach*. Our work so far can be seen as early work towards a true JSR2 reconstruction, and our next steps in this research direction will be dedicated to finding such a joint problem formulation.

8.2 Upcoming and Future Steps

8.2.1 Computer-aided versus manual diagnosis pipeline

Given the sequence of proposed (semi-)automatic methods for midbrain and SNE segmentation, as well as for computer-aided PD diagnosis, we are still missing one important experiment. Thus, one of our immediate next steps will be to *evaluate the whole pipeline of methods against manual expert-segmentations* and compare the diagnostic outcome.

One hypothesis is that the computer-guided midbrain segmentation and SNE detection methods do not only speed up the procedure, but also increase objectivity. We plan to learn our models from segmentations from several human raters, for which we are currently obtaining a third human-performed segmentation of our 22-subject dataset. Given training by three people instead of one (as used in this thesis), we expect the model to improve its generalization behaviour, while hopefully performing some sort of *consensus segmentation/detection of the midbrain* region.

For evaluation, we plan to compare the automated segmentation performances of our methods against a *consensus opinion* formed from the three manual segmentations of our human observers. In order to find the "true" segmentation underlying the three manual opinions, we plan to use an algorithm for "*Simultaneous truth and performance level estimation (STAPLE)*", proposed by Warfield, Zou and Wells in [172].

Eventually, the final test of our computer-aided segmentations will be whether they *produce similar diagnostic accuracy* as when human segmentations are used. Possibly, due to improved objectivity and consensus behaviour, the diagnostic *accuracy might even improve*. Naturally, we plan to validate this hypothesis thoroughly, which is why we are currently extending our dataset, from the 22-subject databse to a larger dataset of around 60 people, including multi-modal image information.

It is important to note that even with computer-aided "diagnosis" methods, the actual diagnosis of PD cannot be made with 3D-TCUS alone. As seen in the diagram of Brooks (see figure 1.1), TCUS is only a small building block of the complex task of PD diagnosis, which will always rely partly on other medical imaging methods, but particularly on detailed neurological examination. However, especially the prospect of *early PD diagnosis* and the usage of *TCUS as a screening method* remains appealing and we thus hope to have laid some early groundwork for a possible routine usage of TCUS with our methods in future.

8.2.2 Upcoming Multi-Modal US-MRI Recordings

The dataset used in this thesis comprises 22 subjects, 11 healthy controls and 11 diagnosed PD patients. While we were able to demonstrate many of our objectives on this dataset such as computer-aided diagnosis and automated segmentation, it has several flaws. We see four main shortcomings of our dataset, which are enlisted in the following. We also make suggestions on improving them as follows:

- **Matched demographics:** The two groups were not age-matched (average age). As reported in tables 4.1 and 4.2, the average age of healhty controls was 55.6 years, while the average age of PD patients was more than ten years older at 65.8 years. This might have partly influenced our results so far, but more importantly, we cannot use demographic data in our SVM-based classifier for computer-aided diagnosis. In order to make this analysis, we need to acquire data on a dataset with comparable demographics.

- **Increased cohort size:** The cohort size is not large enough for solid data analysis and testing of certain hypotheses we have. For example, it would be interesting to investigate whether the spatial distribution of SNEs within the midbrain correlates to PD sub-types such as tremor-dominant, hypokinetic-rigid or aequivalent-type (see chapter 1.1).

- **Recording of RF data:** We have performed all our analyses using 3D B-mode ultrasound data. As we mentioned in chapter 6.2.1 concerning related work on lesion detection in 3DUS, it would be highly interesting to perform SNE analysis using raw radio-frequency (RF) ultrasound data. This notion is also backed up by Noble [109], who reports on superior performance of tissue characterization methods if RF data is used.

- **Multi-modal data with MRI:** As mentioned, the dataset we used is uni-modal, i.e. it only contains 3D B-mode ultrasound data. Especially during our work towards improved reconstruction using the JSR2 algorithm (chapter 7), we realized that it is difficult to estimate the linear and non-linear errors of our data acquisition and reconstruction. On the other hand, if we propose an advanced reconstruction method such as JSR2, we need another modality like MRI for validation.

Objective: According to these four main shortcomings, we have started recording a new dataset, with the following features, which aim to remedy the above-mentioned shortcomings :

1. Matched demographics between the healthy control and PD groups,

2. Larger cohort size, i.e. approximately 60 people, with 30 healthy controls and 30 PD patients, ten each from the groups tremor-dominant, hypokinetic-rigid and aequivalent type,

3. Recording of 3D RF data *and* 3D B-mode data,

4. Acquisition of optically co-registered multi-modal MRI data with several sequences (T1, T2 etc.).

Methodology: We have already started the acquisition of this dataset. We will briefly explain the technical methodology in the following. Since our recordings are at an early stage and we plan to write more detailed publications on this dataset in future, we keep explanations brief at this point.

RF and B-mode data are recorded in 3D. We use almost the identical acquisition setup as we did so far (see chapter 3.1). On each side of the bone windows, we first acquire B-mode data, then RF data, before we switch to the next side. For RF data acquisition, we use a Sonix MDP ultrasound machine (Ultrasonix, British Columbia, Canada) with Ulterius SDK [51] and a custom-written recording software. The 3D B-mode and RF data is optically co-registered through the forehead reference target we proposed and used already in chapter 3.1.

Each subject also undergoes a multi-modal MRI acquisition session immediately prior to, or after 3DUS acquisition. Since the cranial gradient-coil is very tightly fitting to the skull during MRI scans, we cannot attach markers rigidly enough for co-registration of MRI to US. Instead, we use a surface-registration method. For this, we sample the "skull" or rather skin surface of the subject using an optically-tracked pointer device, which we carefully slide along the subject's skin and face, with preference to bony regions to avoid indentation of the skin. All skin surface points are recorded within the coordinate system of the forehead reference target - the same coordinate system in which we record our ultrasound image frames. From the MRI volume, we obtain the skin and face surface through a simple segmentation method based on thresholding and morphological operations.

Both skin surface regions can be registered using any point-cloud registration methods. We use the Iterative Closest Points (ICP) method [26], after a rough manual alignment of body axes between the US and MRI scans.

First Results: At the current stage, we have already acquired data of around 20 subjects. We also already reconstructed a few of them, and we present some first and preliminary results from one of those subjects.

Figure 8.1 shows the two point clouds, i.e. a subset of the cranial skin surface sampled in US space (panel A), the whole-cranium skin surface segmented from MRI (panel B) and their co-registration after ICP.

The result of the co-registration is that we can transform the 3DUS volume into MRI space, where further analyses can be performed, or where algorithms like JSR2 can be validated. Figure 8.2 shows one exemplary overlay. We have not performed a quantitative study yet, where we assess registration quality e.g. in terms of target registration error (TRE). However, qualitatively, the overlay looks plausible.

For example, in the top left image of figure 8.2, it can be seen well that the ultrasound echoes from the opposite cranial walls are well co-registered with the bone regions in MRI. Also, the midbrain is partly well-registered (see for example the middle panel in figure 8.2, where US echoes correlate well with the MRI midbrain).

Slight mis-registrations exist, however. This may be due to insufficient surface registration, which could be remedied by using more robust methods than ICP. On the other hand, these mis-registration could be an indicator for the non-linear physical errors due to ultrasound scanning through the skull bone. It is these mis-registration that motivate advanced reconstruction algorithms like our JSR2 approach. In any case, our preliminary results motivate us to

further pursue our 3DUS B-mode and RF data recordings for this novel, larger and improved dataset.

8.2.3 Multi-center study

A more long-range goal of our study is to motivate other research groups in the TCUS community to also acquire 3D-TCUS data. While individual groups have performed research to certain aspects of TCUS (see chapter 1.2.1), the larger multi-center studies [24, 99, 22] were able to be more conclusive. In particular, multi-center studies are useful for assessment of the objectivity of the method, as well as inter-rater observabilities [99, 24, 162].

While the availability of 3D-TCUS is still limited at the moment, one long-range goal of our efforts is to attract other researchers for collaborations, or motivate them to upgrade to 3DUS as well, allowing us to conduct better analyses of the 3DUS method across institutes.

8.2.4 Extension to other neurological movement disorders

As we have briefly discussed in section 1.2.2, TCUS is not only useful for differential and early diagnosis of Parkinson's disease (PD), but also for diagnosis and discrimination of other neurological neurological movement and neuropsychiatric disorders. Next to varying forms of PD, these include for example corticobasal degeneration, dementia, depression or dystonia.

In this thesis, we have only applied our 3D-TCUS methods for PD diagnosis. In future, we would also like to extend our analyses to other neurological movement disorders. The benefit is that our setup and data acquisition approach would need no modifications, due to the flexibility of 3D Freehand ultrasound. However, we would need to acquire further study data with patient cohorts diagnosed with respective diseases. Furthermore, we would need to develop custom segmentation methods. The midbrain segmentation method can be extended for multi-region detection, by adapting cost-function and optimization routines. The Random Forest approach, however, is inherently capable of multi-region detection and segmentation and could be readily applied to multi-class data. Using such adapted segmentation methods, we could exploit the potential of TCUS further, and contribute to current studies by making use of 3D acquisition and data anlysis.

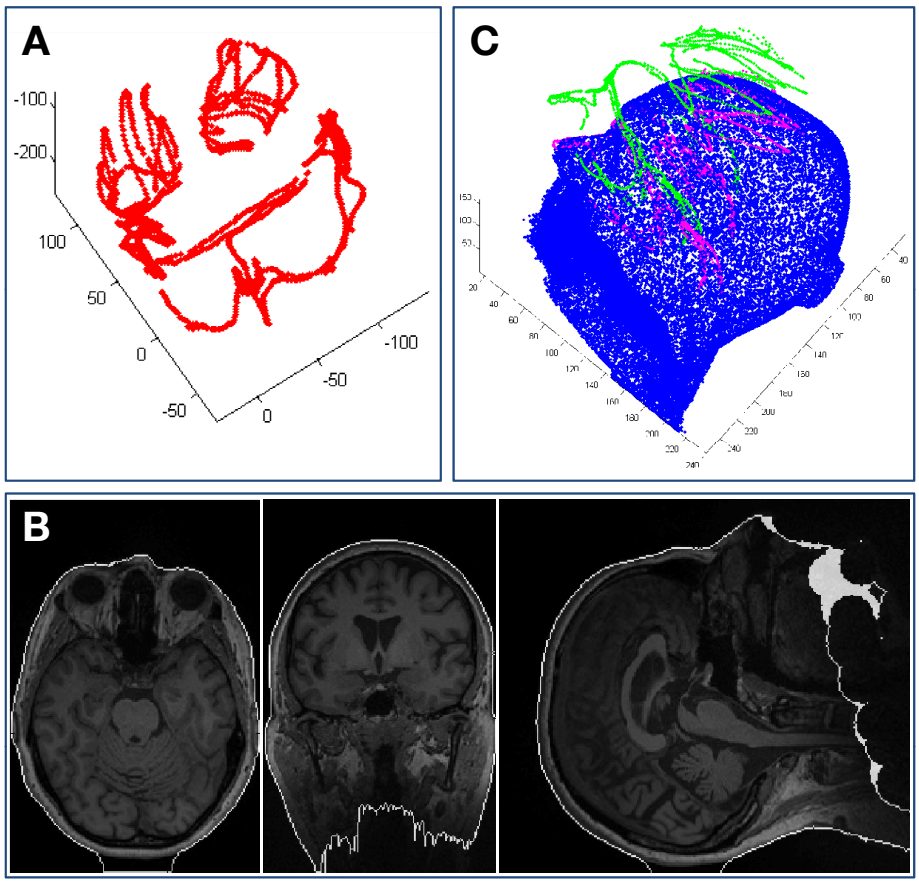

Figure 8.1: First result from our new multi-modal dataset acquisitions. We obtain 3DUS data (B-mode and RF), as well as different modes of MRI scans. We co-register both datasets optically, by sampling the skin surface in US space (A), as well as roughly segmenting the skin surface in MRI space (bright outlines) (B). The registration of both point clouds is performed e.g. using ICP [26] (panel C, blue point cloud = MRI, green = US skin surface prior to ICP, magenta = US matched to MRI using ICP).

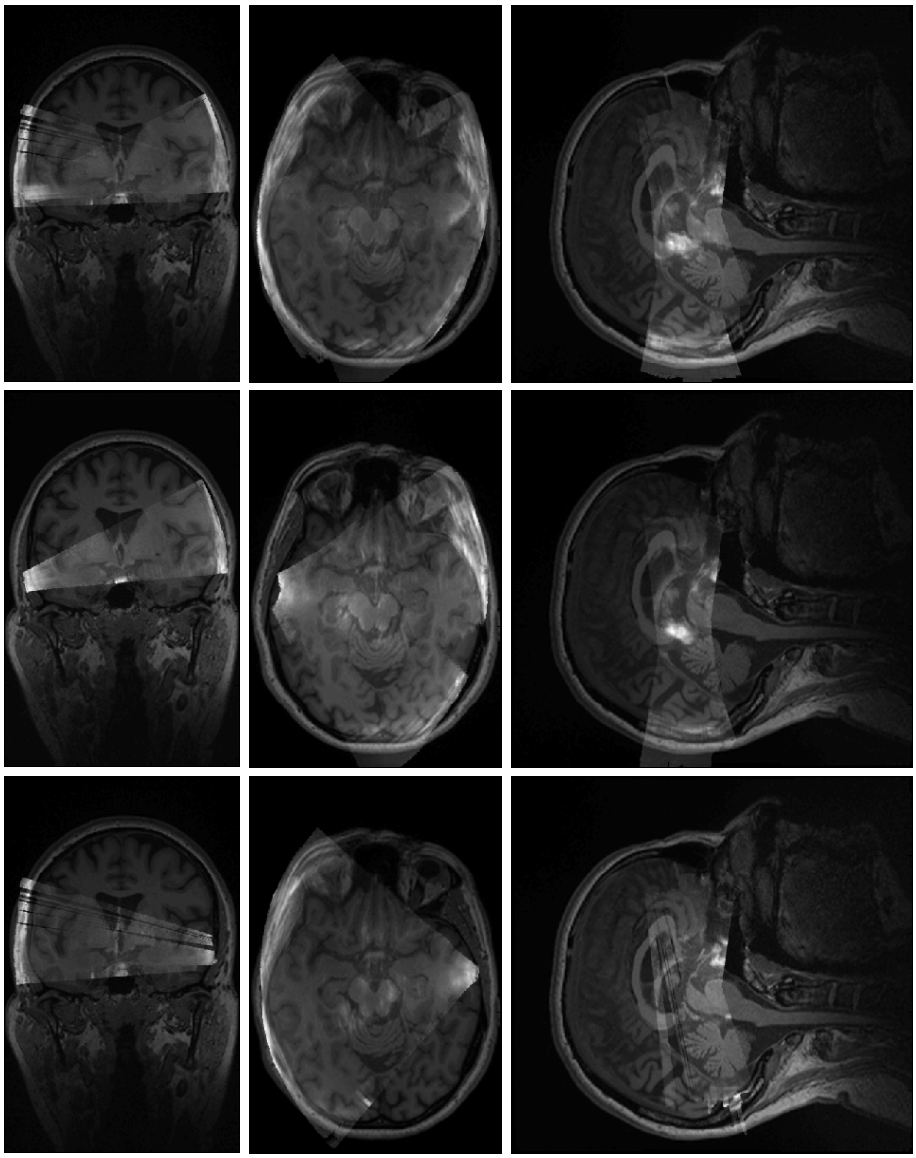

Figure 8.2: First result from our multi-modal acquisition (one subject shown), with coronal, axial and sagittal views (left, middle, right column). The top row shows an overlay of 3D B-mode US and MRI, with bi-lateral reconstruction. The middle and bottom row show uni-lateral reconstructions from left and right bone window, respectively. Cranial walls seem well-matched between US and MRI (top left), as well as the midbrain (middle, middle).

9 Conclusion

In the past decades, transcranial ultrasound (TCUS) has been established as a non-invasive, quick and cheap method for differential and early diagnosis of Parkinson's disease (PD). Apart from PD, TCUS can also be used for diagnosis or discrimination of other neurological movement and neuropsychiatric disorders, such as dystonia, dementia or depression. However, despite the potential of the technique, there has also been criticism, mainly concerning the subjectivity of the method and the necessity for sonographers with high experience, in order to maintain high diagnostic reliability.

In this thesis, for the first time, we have proposed *and* demonstrated the extension of the method from 2DUS to 3DUS, with the goals of increased objectivity, easier image acquisition and possibly advanced diagnostic value, due to the analysis of hyper- and hypo-echogenicities in 3D. We have shown quantitative results from an unprecedented 3D-TCUS study on a study cohort with 11 healthy controls and 11 PD patients, demonstrating sensitivities and specificities for PD diagnosis which are comparable to the state-of-the-art.

Furthermore, we have taken first steps in making the complex data analysis in 3D more clinically applicable, by introducing *a first pipeline of methods for computer-aided PD diagnosis based on 3D-TCUS*. To this end, we have introduced several novel techniques for 3D segmentation of midbrain and substantia nigra (SN) hyper-echogenicities, which are adapted to physical properties of US imaging. These methods incorporate several anatomic priors, such as voxel intensities, surface shapes and spatial distribution of speckle patches, in order to increase their robustness in light of the challenging properties of 3D-TCUS images. We have also proposed a first step towards advanced reconstruction of the midbrain area, by making use of the bi-lateral, multi-view acquisition setup which we have introduced in TCUS for the first time.

We hope that with our contributions, we have laid some groundwork in the area of 3D-TCUS, allowing other research groups to follow up and further investigate this interesting and motivating research area. Most importantly, however, we believe that we have contributed significantly towards computer-aided diagnosis of Parkinson's disease, possibly improving the chance for early diagnosis and onset of therapy for affected patients in future.

A

ROBOCAST and ACTIVE

This thesis was funded by two European projects under the "Seventh Framework Programme" (FP7) called ROBOCAST (FP7-ICT-2007-215190) and ACTIVE (FP7-ICT-2009-6-270460). Both projects deal with robot-assisted neurosurgery. While ROBOCAST was aimed at minimally-invasive keyhole neurosurgery, ACTIVE is more focused on ensuring active constraints for the surgical robots during open-craniotomy epilepsy surgery.

In ROBOCAST, our contributions were two-fold:

1. As work package leader, we supervised and managed the fourth work-package, "WP4 - Planning, navigation and advanced visualization", dealing with the design and development of a pre-operative planning software, intra-operative plan-update due to tissue shifts, as well as error measurement and propagation for on-line estimation and adaptive visualization of planning path as well as tracking and navigation accuracy. Overall, the work package aimed at testing and validating the proposed integrated surgical navigation solution for the entire project.

2. In our main project task, our aim was to design, develop and integrate a 3D (Freehand) Ultrasound system for image guided, minimally-invasive neurosurgery. We developed and integrated a similar system setup and calibration method as described in chapter 3.1, however without the need for bi-lateral acquisition.

In ACTIVE, our aim was to compute brain shift and brain deformation using 3D ultrasound images, such that the pre-operative plan based on MRI could be updated and robotic behaviour could be customized using intra-operative data. We proposed a two-stage approach, based on (1) a rigid registration of pre-operative MRI to 3DUS using segmentation and feature-based registration of equivalent anatomic structures in both modalities and (2) sequential brain shift calculations through deformable 3DUS-3DUS registrations intermittently during surgery, whenever the surgical workflow allows. The approach is summarized in figure A.1. In the following sections of Appendix A, we will briefly cite three paper contributions by the thesis author.

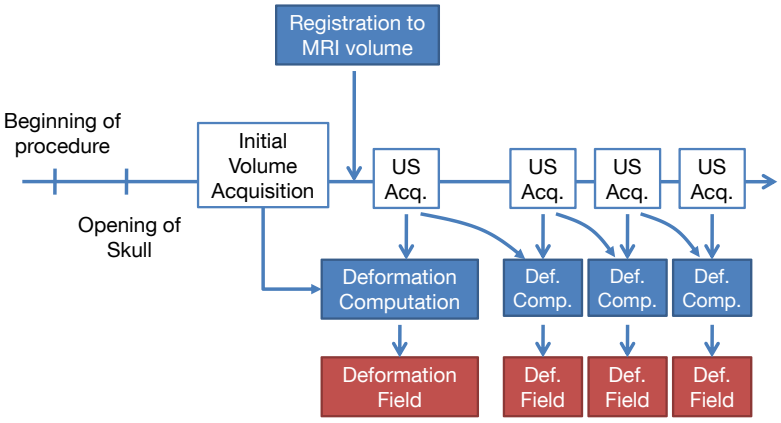

Figure A.1: Overall US-MRI registration approach and registration workflow in ACTIVE.

A.1 Advanced Planning and Intra-operative Validation for Robot-Assisted Keyhole Neurosurgery In ROBOCAST

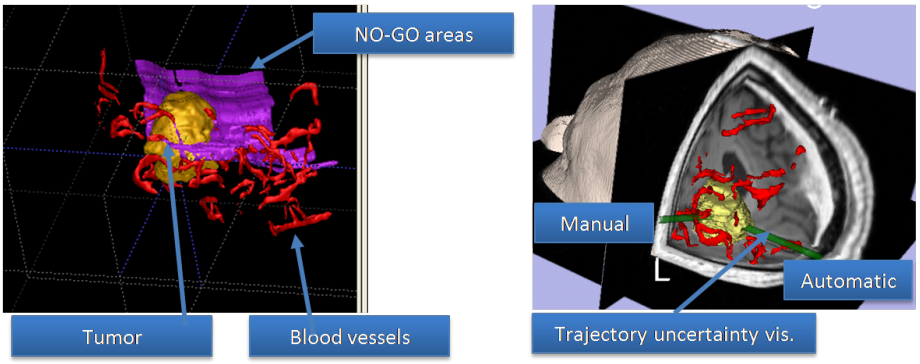

Figure A.2: Screenshot excerpts from our pre-operative planning modules. The left part image shows the definition of a target (tumor), geometric constraints for the trajectory in form of no-go areas and critical structures such as blood vessels combined into a crude risk atlas. The right image shows two automatically proposed trajectories with minimal risk (i.e. maximal distance to non-allowed regions). The incorporation of functional regions, fibres and statistical multi-patient anatomy into the risk atlas is currently under development.

Seyed-Ahmad Ahmadi, Tassilo Klein, Nassir Navab, Ran Roth, Reuben R Shamir, Leo Joskowicz, Elena DeMomi, Giancarlo Ferrigno, Luca Antiga, Roberto Israel Foroni [3]. ROBOCAST is a multi-national project comprising

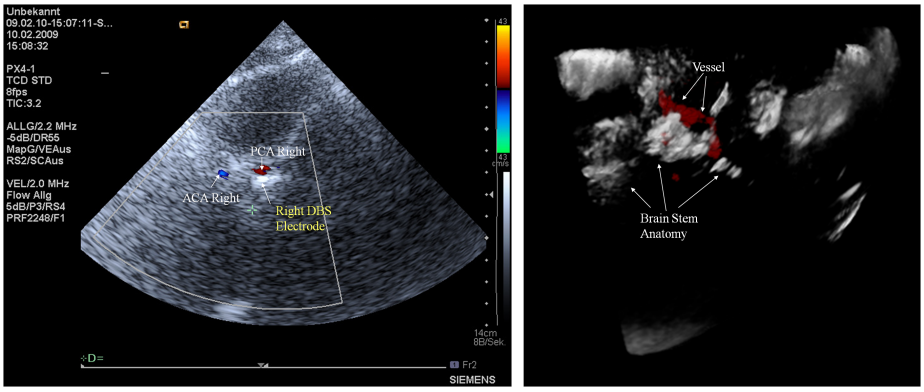

Figure A.3: Transcranial ultrasound scans of the basal ganglia in 2D on patient 1 (left) and in 3D on patient 2 (right). The cross-section of the DBS electrode in relation to the PCA (arteria cerebri posterior) can be clearly seen in patient 1. The right image shows a 3D reconstruction of the brain stem anatomy of patient 2 (white) and the reconstructed PCA vessel (red).

several institutes which aim at outlining and implementing a prototype system for advanced, robot-assisted keyhole neurosurgery.

This paper reports mainly on software and sensor aspects of the system in the pre- and intra-operative stage. We describe a comprehensive workflow of planning steps provided to the surgeon in a wizard-like manner. As a novelty, we present a method for automatic trajectory planning based on a statistical and patient-specific risk atlas. Intra-operative monitoring of system uncertainty is proposed. Furthermore, the usage of 2D and 3D Freehand Ultrasound for intra-operative validation is motivated and theoretically outlined.

After the first of three years project runtime, we present current work in progress and preliminary results on several patient studies concerning automatic path planning, trajectory localization errors as well as ultrasound imaging on one, six and three patients, respectively. The results underline the usefulness and significance of proposed methods, both within the scope of the ROBOCAST project as well as for conventional keyhole neurosurgery.

A.2 User friendly graphical user interface for workflow management during navigated robotic-assisted keyhole neurosurgery

Seyed-Ahmad Ahmadi Ahmadi, Francesco Pisana, Elena DeMomi, Nassir Navab, Giancarlo Ferrigno [5]. Minimally-invasive neurosurgery requires careful pre-operative planning of trajectories prior to the execution of the surgery. This work was developed within the EU project ROBOCAST, which aims at developing an integrated solution for advanced planning and intra-operative navigation of robot-assisted neurosurgical procedures. A combination of three robots offering a large working volume with high local precision and accuracy, together with an innovative flexible probe allows for research towards novel surgical procedures such as Multi-Target Treatment procedures. In the recent years, the research community dealing with intra-operative navigation systems has identified the need for analysis of surgical workflow and context-sensitive user interfaces in the OR. In this work, we present our work-in-progress on a user-friendly user interface, which intra-operatively guides the surgeon through the execution of the pre-operative plan in form of a sequential workflow wizard. Each step of the workflow spawns a window with clear visual and verbal instructions while a state-machine in the background controls whether the correct sequence of actions is performed and which workflow transitions are allowed. Therefore, the surgeon is optimally guided through the workflow, allowing him to focus on the medical procedure at hand. The integration of further components from our overall system into the Touchscreen user interface guarantees that user input is correctly propagated within the system and that the current workflow step and surgical activity drive the internal behavior of the system, making our system partly context-aware.

A.3 Rigid US-MRI Registration Through Segmentation of Equivalent Anatomic Structures

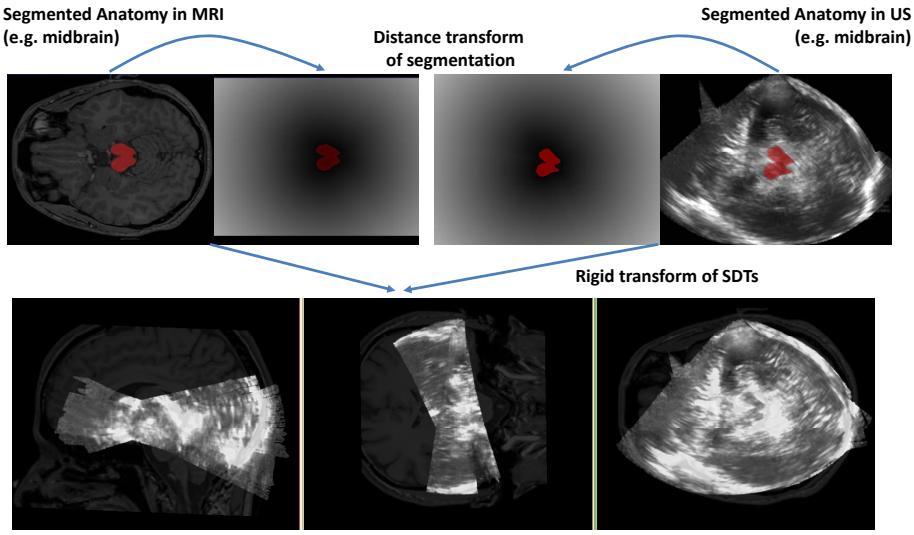

Figure A.4: Illustration of the rigid registration approach for T1-MRI (top left) and 3D-TCUS (top right), using signed distance transforms (SDT) of the segmented surfaces. The bottom row shows an example registration result.

Seyed-Ahmad Ahmadi, Tassilo Klein, Annika Plate, Kai Boetzel, Nassir Navab [4]. Multi-modal registration between 3D ultrasound (US) and magnetic resonance imaging (MRI) is motivated by aims such as image fusion for improved diagnostics or intra-operative evaluation of brain shift. In this work, we present a rigid region-based registration approach between MRI and 3D-US based on the segmentation of equivalent anatomic structures in both modalities. Our feasibility study is performed using segmentations of the midbrain in both MRI and 3D transcranial ultrasound. Segmentation of MRI is based on deformable atlas registration while for 3D US segmentation, we recently proposed an accurate and robust method based on statistical shape modeling and a discrete and localized active surface segmentation framework. The multi-modal registration is performed through intensity-based rigid registration of signed distance transforms of both segmentations. Qualitative results and a demonstration of the basic feasibility of the region-based registration are demonstrated on a pair of MRI and challenging 3D transcranial US data volumes from the same subject.

B

Other Contributions

B.1 Automatic Segmentation of the Carotid Artery in freehand Ultrasound

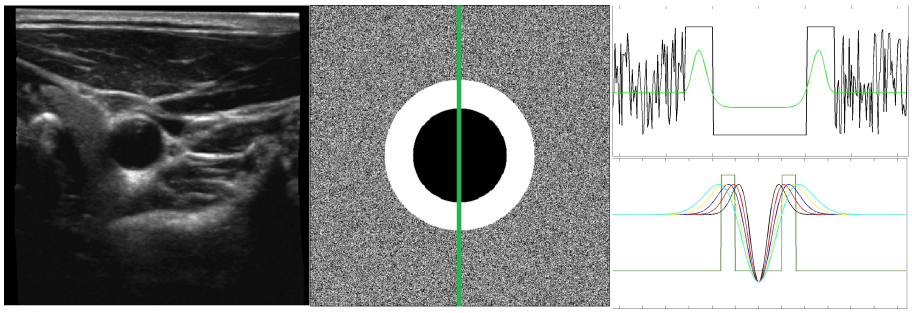

Figure B.1: Left: Common carotid artery cross-section; Middle: Ideal lumen cross-section with line segment overlaid; Right: Line segment of ideal lumen cross-section with $\frac{\partial^2}{\partial k^2} kernel$ overlaid: proposed kernel (top, in green), multiple attempts to fit the gaussian kernel (bottom; noise removed for better visualization)

Paulo Waelkens, Seyed-Ahmad Ahmadi, Nassir Navab [166]. We propose a Hessian matrix based multiscale tubular structure detection (TSD) algorithm adapted to 3D B-mode vascular US images. The algorithm is designed to highlight blood vessel centerline points and yield an estimate of the cross-section radius at each centerline point. It can be combined with a simple centerline extraction scheme, yielding precise, fast and fully automatic lumen segmentation initializations.
TSD algorithms designed with CTA and MRA datasets in mind, e.g. the Frangi Filter [57], are not capable of reliably distinguishing centerline points from other points in vascular US datasets, since some assumptions underlying these

algorithms are not reasonable for US datasets. The algorithm we propose, does not have these shortcomings and performs significantly better on vascular US datasets.
We propose a statistic to evaluate how well a TSD algorithm is able to distinguish centerline points from other points. Based on this statistic, we compare the Frangi Filter to various versions of our new algorithm, on 11 3D US carotid datasets.

B.2 Ultrasound Bone Detection Using Patient-Specific CT Prior

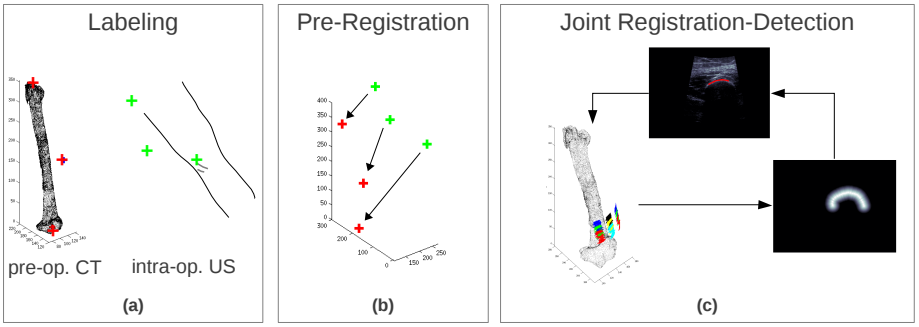

Figure B.2: Overview of the proposed workflow, (a) shows the labeling of the correspondence points in both modalities, (b) the pre-alignment towards point-based registration and (c) the combined registration and bone detection approach in ultrasound.

Julian Beitzel, Seyed-Ahmad Ahmadi, Athanasios Karamalis, Wolfgang Wein and Nassir Navab [18]. Registration of pre-operative CT datasets to intra-operative 3D freehand ultrasound has been of high interest for computer assisted orthopedic surgery. Feature-based registration relies on an accurate detection of the bone surface in the B-mode ultrasound images. In this work we present a fully automatic bone detection approach for US. The pre-operative CT is utilized to create a patient-specific bone model for our joint detection-registration framework. The model provides a geometric constraint for accurate and robust detection. Simultaneously to the detection, our method yields a close estimate of the rigid transformation from US to CT, which can be used as an initialization for further refinement through sophisticated intensity-/feature-based registration methods. We evaluated our approach on datasets of the human femur acquired in a cadaver study and demonstrate a mean bone detection error of below $0.4mm$.

B.3 Multi-modal Image Registration for pre-operative planning and intra-operative guidance using Adaptive Distance Measures

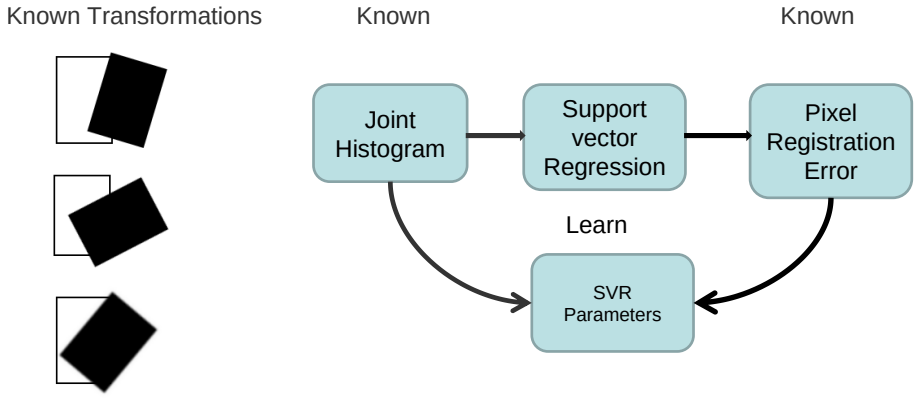

Figure B.3: Schematic of the training phase in our machine learning-based regression algorithm. Having a pair of manually registered images, Q known rigid transformation matrices are individually applied on the source image. Then the calculated joint histogram feature vector and PRE labels are given to the support vector regressor in order to learn the relationship between them.

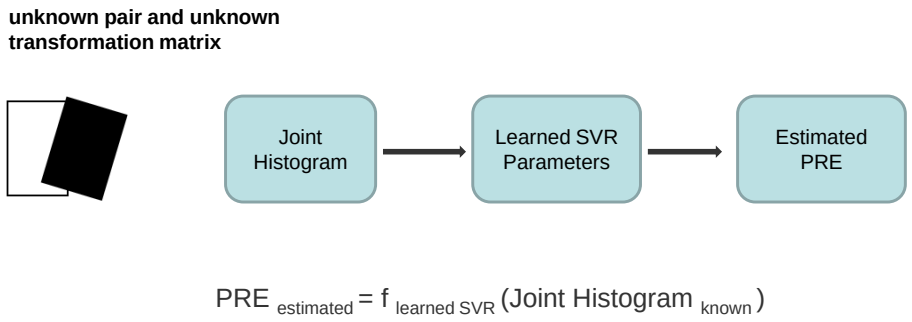

Figure B.4: TThis figure demonstrates the test procedure which assures us to have a convex regression function. First a set of transformations are applied on a pair of images. This image pair can be different from the pair which was used in the training phase. Calculated joint histograms are then given to the learned SVR structure in order to estimate the correspondent PRE value.

Banafsheh Jalali, Master thesis [76]. Modern medicine often employs various imaging modalities such as CT, MRI, PET or ultrasound for improved diagnosis, pre-operative planning, intra-operative navigation and post-operative monitoring. In order to track anatomic changes in one modality over time or in order to fuse complementary information from different modalities, automatic

methods for data fusion or registration are required. A particular problem lies in the image fusion across different modalities, since their underlying data often stands in a non-linear relationship to each other. Research of the past decade has proposed a large number of solutions for multi-modal image registration. Often, pairs of modalities to be registered are chosen according to the target organ and type of the surgical operation. In this thesis, 3D rigid registration of CT to MRI and 2D registration of ultrasound to MRI have been covered.

While mutual information (MI) has become the established method for performing CT to MRI registration, current proposed methods are not successful to fully cover all the aspects of ultrasound to MRI registration problem. Some of the proposed solutions to this problem are limited to a specific target organ, some others do not have the required accuracy to be utilized for specific organs, while others suffer from being highly computational expensive.

In this thesis we are aiming at learning an adaptive metric for performing rigid registration between any two pairs of imaging modalities. The goal is to learn the relationship between two modalities to be registered from a pair of those modalities captured before the surgery and use this as prior information, e.g. during intra-operative registration. Our method is based on previous work for performing 2D rigid registration utilizing a machine learning algorithm. In order to calculate the exact motion of the tissue during surgery, 3D information is required. Therefore, we have investigated the possibility of extending the idea into 3D scenarios. Then the method is evaluated for the more difficult rigid registration problem of ultrasound to MRI in 2D.

B.4 A user-friendly, clinical-usable freehand ultrasound probe calibration for clinical practicability

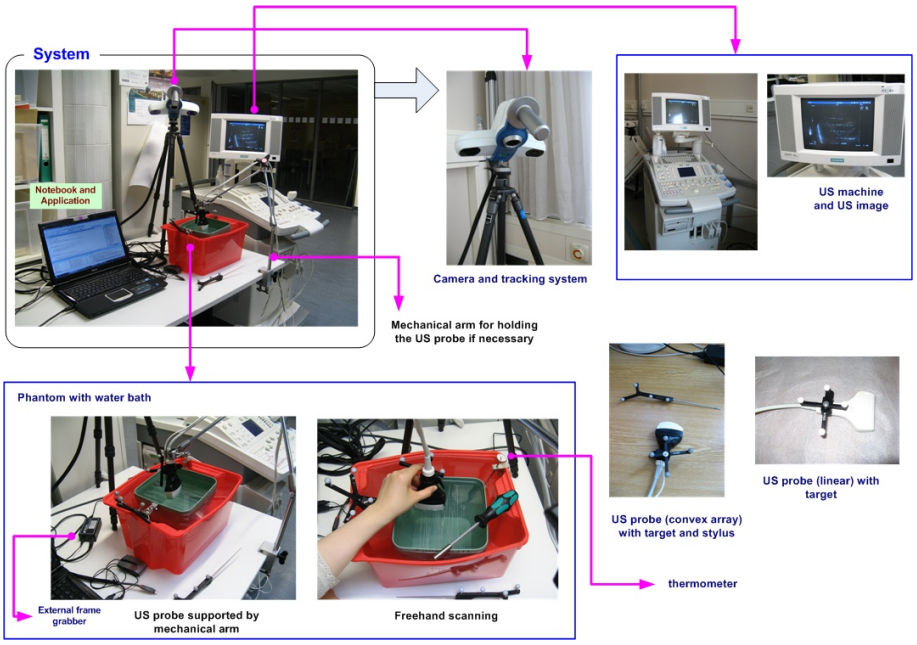

Figure B.5: The US calibration experiment setup.

Mei Chuan Chen, Master thesis [38]. Ultrasound (US) is a relatively inexpensive, safe, and non-invasive imaging technology. It has been used for clinical applications in fields such as intra-operative imaging by obtaining a 3D anatomy model, or multi-modality registration (US with CT or MR data). It is hence necessary to perform a reliable US probe calibration to get the accurate positions of the US images. Furthermore, rapidness and efficiency are the necessary factors to be considered when performing a clinically applicable calibration.

In this project, a simple Z-phantom made up of 11 Z-shaped structures that are distributed across five layers is used to perform the probe calibration. An automatic and robust feature segmentation method has been developed to facilitate the US calibration procedure. This method does not require additional support from the phantom construction and can be applied on the US images with more artifacts.

An application providing an user-interface has been developed to perform the US calibration. It only takes one mouse-clicking for one frame to capture the frames needed for a calibration. The calibration results show that only a few frames are needed to get a good calibration. Therefore, it only takes less

than one minute to perform a calibration if the pixel scales have been calibrated. Thanks to the automatic segmentation and freehand scanning, it enables a rapid procedure of the calibration and is clinically applicable. According to the study of the accuracy and precision of the calibration, a precise calibration has been guaranteed and an accurate calibration can be achieved if more valid frames are used.

B.5 Tissue-mimicking multi-modal brain phantom

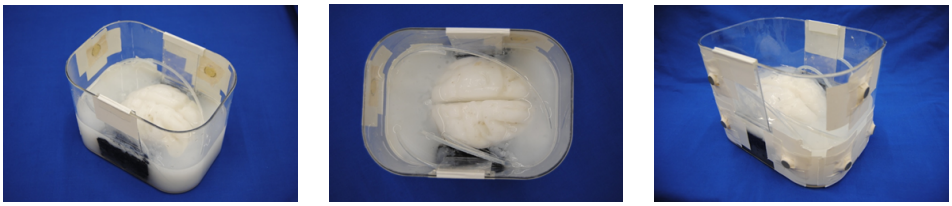

Figure B.6: Three views on the final brain phantom, note the arterial/vein tubes for in- and output of the vessel system and one additional for the balloon, the black foamed plastic strips at the side suppress the buoyancy of the phantom if water comes below the PVA-fundament. The infrared-markers at the outside of the shell are for the 3D-ultrasound registration.

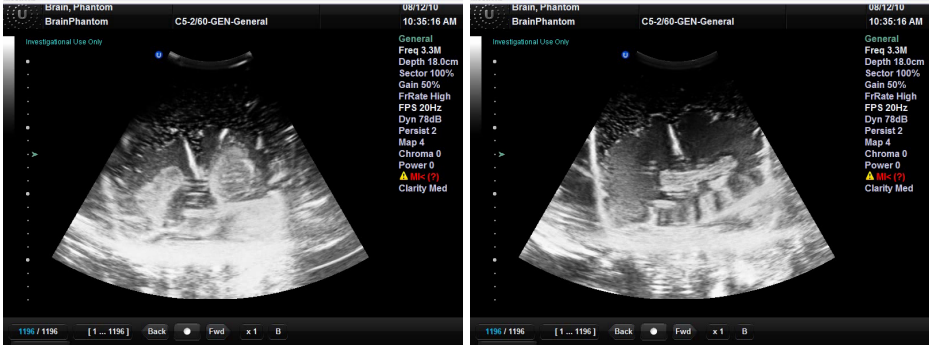

Figure B.7: The **B-Mode Ultrasound** images show parts of the internal structure, the left image shows a slice through the cerebellum, the right displays the pons in the middle, furthermore the interbrain, midbrain and the balloon.

Julian Beitzel, Interdisciplinary Project Tech Report [17]. A computer-assisted analysis for acquired medical images is highly desirable in order to deal with the increased complexity of 3D imaging, in particular in case of 3D transcranial ultrasound examinations. However, in order to create reliable computer-aided diagnostic tools, qualitative and quantitative testing and validating are of high priority due to the serious consequences that can occur from a mislead diagnosis. Three common ways to accomplish that are:

1. Creation of artificial reference data, i.e. creating artificial (e.g. digital) data to serve the purpose of early experiments and feasibility studies but generally lacks the complex features of realistic data.

2. Creation of a phantom that mimics the desired target anatomy, e.g. in spatial dimensions, physical properties, appearance etc., for quantitative analysis and realistic imaging.

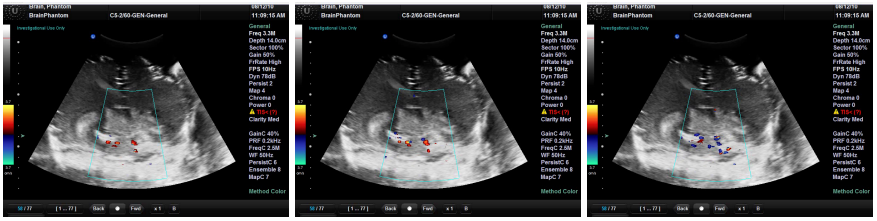

Figure B.8: The **Doppler Ultrasound** shows the flow of water through the tube system, the color indicates the intensity and the direction.

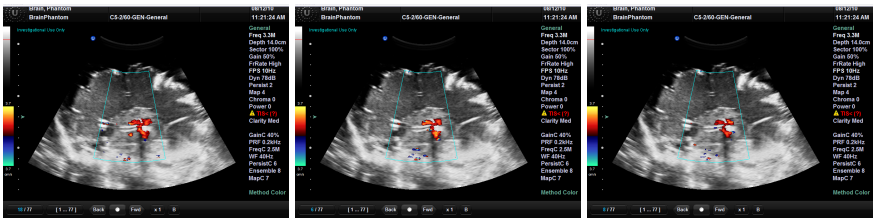

Figure B.9: These **Doppler Ultrasound** images show the signal at the connection points, where the tubes are connected.

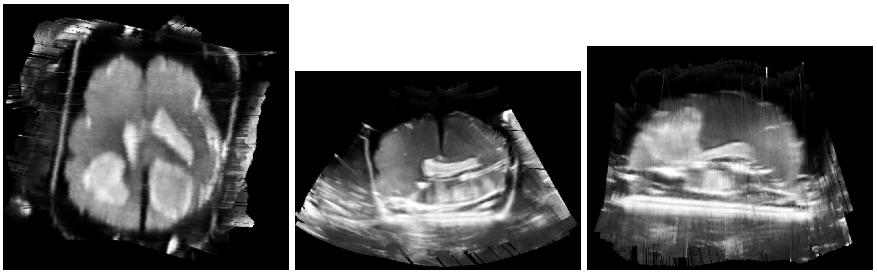

Figure B.10: The **3D Ultrasound** images shows the axial, coronal and sagittal slice through the phantom. Especially the axial slice shows the shift of the internal structures, e.g. the cerebellum, due to the buoyancy of cryogel during the creation process.

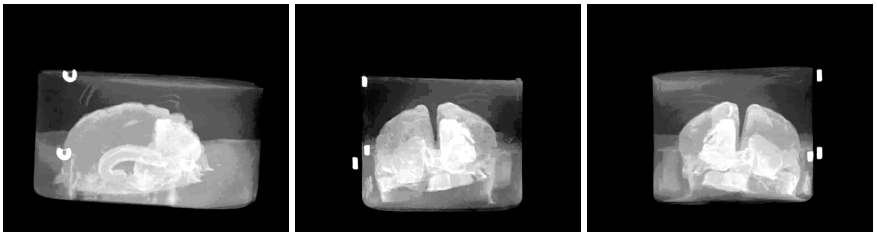

Figure B.11: These T1-weighted images of the **Magnetic Resonance Imaging** show the axial slice on the left, and the back- and front-view in the middle/right. Similar to the 3D-US, the MRI shows the shift of the internal structures in the production process. Note that the bright spots around the phantom are the reflecting gadolinium markers.

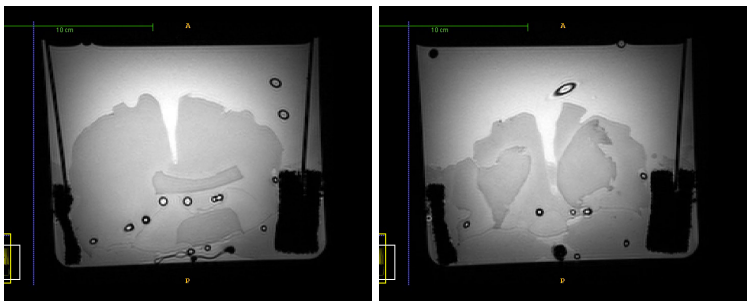

Figure B.12: The **T2-weighted MRI** images clearly shows the tubes inside the phantom, furthermore the pons in the middle of the left image, below the interbrain and the midbrain at the bottom. The right image shows the cerebellum and the balloon beneath.

3. Ex-vivo studies using human or animal tissue, e.g. a liver or brain for achieving maximally realistic tissue properties but with limited lifetime of the sample (e.g. due to decay).

This report focuses on the creation of a multi-modal and -functional, brain mimicking phantom which can be used for many applications, including linear and deformable multi-modal registration, segmentation, visualization, and navigated surgery simulations, e.g. for simulation of intra-operative brain shift in simulated neurosurgery. The phantom mimicks six distinct anatomic areas of the human brain: (1) cerebrum, (2) interbrain, (3) midbrain, (4) cerebellum, (5) pons and (6) afterbrain. Furthermore, the phantom features a vessel tree which roughly corresponds in structure to actual human brain vasculature. It also features a means of artifical deformation for simulation of brain shift. The selected material is Polyvinyl Alcohol cryogel (PVAc), which has been shown to offer realistic tissue-mimicking properties with favorable properties in terms of material processing.

Bibliography

[1] CGAL, Computational Geometry Algorithms Library. Http://www.cgal.org

[2] Ahmadi, S.A., Baust, M., Karamalis, A., Plate, A., Bötzel, K., Klein, T., Navab, N.: Midbrain Segmentation in Transcranial 3D Ultrasound for Parkinson Diagnosis. In: Proceedings Internation Conference Medical Image Computing and Computer Aided Intervention (MICCAI), pp. 362–369 (2011)

[3] Ahmadi, S.A., Klein, T., Navab, N., Roth, R., Shamir, R., Joskowicz, L., DeMomi, E., Ferrigno, G., Antiga, L., Foroni, R.: Advanced planning and intra-operative validation for robot-assisted keyhole neurosurgery in robocast. In: Advanced Robotics, 2009. ICAR 2009. International Conference on, vol. 1, pp. 1 –7 (2009)

[4] Ahmadi, S.A., Klein, T., Plate, A., Boetzel, K., Navab, N.: Rigid us-mri registration through segmentation of equivalent anatomic structures - a feasibility study using 3d transcranial ultrasound of the midbrain. In: T. Tolxdorff, T.M. Deserno, H. Handels, H.P. Meinzer (eds.) Bildverarbeitung für die Medizin, Informatik Aktuell, pp. 410–415. Springer (2012)

[5] Ahmadi, S.A., Pisana, F., DeMomi, E., Navab, N., Ferrigno, G.: User friendly graphical user interface for workflow management during navigated robotic-assisted keyhole neurosurgery. In: Computer Assisted Radiology (CARS), 24th International Congress and Exhibition. Geneva, CH (2010)

[6] Allott, C.P., Barry, C.D., Pickford, R., Waterton, J.C.: Volumetric assessment of carotid artery bifurcation using freehand-acquired, compound 3D ultrasound. Br J Radiol **72**(855), 289–292 (1999)

[7] Antoniou, A.: Digital Signal Processing: Signals, Systems, and Filters. McGraw-Hill (2005)

[8] Bamber, J.C., Dickinson, R.J.: Ultrasonic B-scanning: a computer simulation. Phys. Med. and Biol. **25**(3), 463 (1980)

[9] Baun, J.: Principles of General and Vascular Sonography, 1st edition edn. ProSono (2004)

[10] Baust, M., Navab, N.: A spherical harmonics shape model for level set segmentation. 11th European Conference on Computer Vision (ECCV) (2010)

[11] Baust, M., Yezzi, A., Unal, G., Navab, N.: A sobolev-type metric for polar active contours. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2011)

[12] Baust, M.T.: Polar Active Contours for Medical Applications. Ph.D. thesis, Technische Universität München, Fakultät für Informatik, Computer Aided Medical Procedures & Augmented Reality (CAMP) / I16 (2012)

[13] Becker, G., Becker, T., Struck, M., Lindner, A., Burzer, K., Retz, W., Bogdahn, U., Beckmann, H.: Reduced echogenicity of brainstem raphe specific to unipolar depression: a transcranial color-coded real-time sonography study. Biol. Psychiatry **38**(3), 180–184 (1995)

[14] Becker, G., Seufert, J., Bogdahn, U., Reichmann, H., Reiners, K.: Degeneration of substantia nigra in chronic parkinson's disease visualized by transcranial color-coded real-time sonography. Neurology **45**(1), 182–184 (1995)

[15] Becker, G., Struck, M., Bogdahn, U., Becker, T.: Echogenicity of the brainstem raphe in patients with major depression. Psychiatry Res **55**(2), 75–84 (1994)

[16] Behnke, S., Berg, D., Naumann, M., Becker, G.: Differentiation of Parkinson's disease and atypical parkinsonian syndromes by transcranial ultrasound. J. Neurol. Neurosurg. Psychiatr. **76**(3), 423–425 (2005)

[17] Beitzel, J.: Design and Development of a multi-modal, multi-purpose brain-mimicking Phantom for 3D Ultrasound Segmentation and Registration Tasks. Tech. rep., Technische Universität München, Fakultät für Informatik (2011)

[18] Beitzel, J., Ahmadi, S.A., Karamalis, A., Wein, W., Navab, N.: Ultrasound bone detection using patient-specific ct prior. In: 34. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). San Diego, USA (2012)

[19] Bereznai, B., Steude, U., Seelos, K., Botzel, K.: Chronic high-frequency globus pallidus internus stimulation in different types of dystonia: a clinical, video, and MRI report of six patients presenting with segmental, cervical, and generalized dystonia. Mov. Disord. **17**(1), 138–144 (2002)

[20] Berg, D.: Marker for a preclinical diagnosis of Parkinson's disease as a basis for neuroprotection. J. Neural Transm. Suppl. **1**(71), 123–132 (2006)

[21] Berg, D., Becker, G., Zeiler, B., Tucha, O., Hofmann, E., Preier, M., Benz, P., Jost, W., Reiners, K., Lange, K.W.: Vulnerability of the nigrostriatal system as detected by transcranial ultrasound. Neurology **53**(5), 1026–1031 (1999)

[22] Berg, D., Godau, J., Walter, U.: Transcranial sonography in movement disorders. Lancet Neurology **7**, 1044–1055 (2008)

[23] Berg, D., Roggendorf, W., Schroder, U., Klein, R., Tatschner, T., Benz, P., Tucha, O., Preier, M., Lange, K.W., Reiners, K., Gerlach, M., Becker, G.: Echogenicity of the substantia nigra: association with increased iron content and marker for susceptibility to nigrostriatal injury. Arch. Neurol. **59**(6), 999–1005 (2002)

[24] Berg, D., Seppi, K., Behnke, S., Liepelt, I., Schweitzer, K., Stockner, H., Wollenweber, F., Gaenslen, A., Mahlknecht, P., Spiegel, J., Godau, J., Huber, H., Srulijes, K., Kiechl, S., Bentele, M., Gasperi, A., Schubert, T., Hiry, T., Probst, M., Schneider, V., Klenk, J., Sawires, M., Willeit, J., Maetzler, W., Fassbender, K., Gasser, T., Poewe, W.: Enlarged substantia nigra hyperechogenicity and risk for Parkinson disease: a 37-month 3-center study of 1847 older persons. Arch. Neurol. **68**, 932–937 (2011)

[25] de Berg, M., Cheong, O., van Kreveld, M., Overmars, M.: Computational Geometry: Algorithms and Applications, 3rd ed. edn. Springer (2008)

[26] Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)

[27] Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)

[28] Blum, T., Heining, S.M., Kutter, O., Navab, N.: Advanced training methods using an augmented reality ultrasound simulator. In: 8th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2009), pp. 177–178. Orlando, USA (2009)

[29] Brechbühler, C., Gerig, G., Kübler, O.: Parametrization of closed surfaces for 3-d shape description. Comput. Vis. Image Underst. **61**(2), 154–170 (1995)

[30] Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)

[31] Brooks, D.J.: Imaging approaches to Parkinson disease. J. Nucl. Med. **51**(4), 596–609 (2010)

[32] Cates, J.: ShapeWorks. URL, http://www.nitrc.org/projects/shapeworks/ (last access 31. Aug. 2012)

[33] Cates, J., Fletcher, P.T., Styner, M., Hazlett, H.C., Whitaker, R.: Particle-based shape analysis of multi-object complexes. In: Proceedings of the 11th international conference on Medical Image Computing and Computer-Assisted Intervention - Part I, MICCAI '08, pp. 477–485. Springer-Verlag, Berlin, Heidelberg (2008)

[34] Cates, J., Meyer, M., Fletcher, T., Whitaker, R.: Entropy-Based Particle Systems for Shape Correspondence. In: X. Pennec, S. Joshi (eds.) 1st MICCAI Workshop on Mathematical Foundations of Computational Anatomy: Geometrical, Statistical and Registration Methods for Modeling Biological Shape Variability, pp. 90–99. Copenhagen, Danemark (2006). URL http://hal.inria.fr/inria-00634198

[35] Chan, T., Vese, L.: Active contours without edges. Image Processing, IEEE Transactions on **10**(2), 266 –277 (2001). DOI 10.1109/83.902291

[36] Chang, F.M., Hsu, K.F., Ko, H.C., Yao, B.L., Chang, C.H., Yu, C.H., Chen, H.Y.: Three-dimensional ultrasound assessment of fetal liver volume in normal pregnancy: a comparison of reproducibility with two-dimensional ultrasound and a search for a volume constant. Ultrasound Med Biol **23**(3), 381–389 (1997)

[37] Chen, L., Seidel, G., Mertins, A.: Multiple feature extraction for early parkinson risk assessment based on transcranial sonography image. In: Image Processing (ICIP), 2010 17th IEEE International Conference on, pp. 2277 –2280 (2010). DOI 10.1109/ICIP.2010.5654216

[38] Chen, M.C.: A user-friendly, clinical-usable freehand ultrasound probe calibration for clinical practicability. Master's thesis, Technische Universität München, Fakultät für Informatik (2010)

[39] Chen, S., Charpiat, G., Radke, R.J.: Converting level set gradients to shape gradients. In: Proceedings of the 11th European Conference on Computer Vision (ECCV 2011) (2010)

[40] Cootes, T., Edwards, G., Taylor, C.: Active appearance models. Pattern Analysis and Machine Intelligence, IEEE Transactions on **23**(6), 681 –685 (2001)

[41] Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding **61**(1), 38–59 (1995)

[42] Crane, K., Llamas, I., Tariq, S.: Real-time simulation and rendering of 3d fluids. In: H. Nguyen (ed.) GPU Gems 3, chap. 30. Addison Wesley Professional (2007)

[43] Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. Int. J. Comput. Vision **72**(2), 195–215 (2007)

[44] Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Tech. rep., Microsoft Research Cambridge, UK (2011)

[45] Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press (2000)

[46] Dalal, N., Triggs, B.: Histogram of Oriented GRadients for Human Detection. In: Proc. Int'l. Conf. Computer Vision and Pattern Recognition (CVPR) (2005)

[47] Davies, R., Twining, C., Cootes, T., Waterton, J., Taylor, C.: A minimum description length approach to statistical shape modeling. Medical Imaging, IEEE Transactions on **21**(5), 525 –537 (2002). DOI 10.1109/TMI.2002.1009388

[48] Davies, R., Twining, C., Taylor, C.: Statistical Models of Shape: Optimisation and Evaluation, 1 edn. Springer Publishing Company, Incorporated (2008)

[49] Delaunay, B.N.: Sur la sphère vide. Bulletin of Academy of Sciences of the USSR (6), 793–800 (1934)

[50] Demirci, S., Lejeune, G., Navab, N.: Hybrid deformable model for aneurysm segmentation. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Boston, Massachusetts, USA (2009)

[51] Dickie, K., Leung, C., Zahiri, R., Pelissier, L.: A flexible research interface for collecting clinical ultrasound images. In: J.K. Udupa, N. Sang, L.G. Nyul, H. Tong (eds.) MIPPR 2009: Multispectral Image Acquisition and Processing, vol. 7494, pp. 749,402+. SPIE (2009)

[52] Engel, K., Toennies, K.D.: Segmentation of the midbrain in transcranial sonographies using a two-component deformable model. Annals of the BMVA **2009**(4), 1–12 (2009)

[53] Fang, Q., Boas, D.A.: Tetrahedral mesh generation from volumetric binary and gray-scale images. In: Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging: From Nano to Macro, ISBI'09, pp. 1142–1145. IEEE Press, Piscataway, NJ, USA (2009). URL http://dl.acm.org/citation.cfm?id=1699872.1700164

[54] Fawcett, T.: An introduction to roc analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)

[55] Floater, M.S., Hormann, K.: Surface parameterization: a tutorial and survey. In: In Advances in Multiresolution for Geometric Modelling, pp. 157–186. Springer (2005)

[56] Foroughi, P., Abolmaesumi, P.: Intra-subject elastic registration of 3D ultrasound images. Medical Image Analysis **10(5)**, 83–90 (2006)

[57] Frangi, A.F.: Three-dimensional model-based analysis of vascular and cardiac images. Ph.D. thesis, Universiteit Utrecht (2001)

[58] Gaenslen, A., Unmuth, B., Godau, J., Liepelt, I., Di Santo, A., Schweitzer, K.J., Gasser, T., Machulla, H.J., Reimold, M., Marek, K., Berg, D.: The specificity and sensitivity of transcranial ultrasound in the differential diagnosis of Parkinson's disease: a prospective blinded study. Lancet Neurol **7**, 417–424 (2008)

[59] Gao, H., Choi, H.F., Claus, P., Boonen, S., Jaecques, S., Van Lenthe, G.H., Van der Perre, G., Lauriks, W., D'hooge, J.: A fast convolution-based methodology to simulate 2-D/3-D cardiac ultrasound images. IEEE Trans Ultrason Ferroelectr Freq Control **56**(2), 404–409 (2009)

[60] Gao, H., D'hooge, J., Hergum, T., Torp, H.: Comparison of the performance of different tools for fast simulation of ultrasound data. In: Ultrasonics Symposium, 2008. IUS 2008. IEEE, pp. 1318 –1321 (2008). DOI 10.1109/ULTSYM.2008.0319

[61] Gao, X., Su, Y., Li, X., Tao, D.: A review of active appearance models. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on **40**(2), 145 –158 (2010). DOI 10.1109/TSMCC.2009.2035631

[62] Gee, A., Prager, R., Treece, G., Cash, C., Berman, L.: Processing and visualizing three-dimensional ultrasound data. Br J Radiol **77 Spec No 2**, S186–193 (2004)

[63] Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for ms lesion segmentation in multi-channel mr images. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2010)

[64] Gotsman, C., Gu, X., Sheffer, A.: Fundamentals of spherical parameterization for 3d meshes. ACM Trans. Graph. **22**(3), 358–363 (2003). DOI 10.1145/882262.882276. URL http://doi.acm.org/10.1145/882262.882276

[65] Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology **143**(1), 29–36 (1982). URL http://radiology.rsna.org/content/143/1/29.abstract

[66] Haralick, R., Shapiro, L.: Computer and Robot Vision, ed. 1 edn. Addison-Wesley Publishing Company (1992)

[67] Hedrick, W., Hykes, D., Starchman, D.: Ultrasound physics and instrumentation, 4th edition edn. Mosby (2004). URL http://books.google.de/books?id=0JZrAAAAMAAJ

[68] Heimann, T., Meinzer, H.P.: Statistical shape models for 3d medical image segmentation: A review. Medical Image Analysis **13**(4), 543 – 563 (2009). DOI 10.1016/j.media.2009.05.004. URL http://www.sciencedirect.com/science/article/pii/S1361841509000425

[69] Hergum, T.: 3D ultrasound for quantitative echocardiography. Ph.D. thesis, Norwegian University of Science and Technology, Faculty of Medicine, Department of Circulation and Medical Imaging, Trondheim, Norway (2009)

[70] Horowski, S., Zettl, U.K., Benecke, R., Walter, U.: Sonographic basal ganglia alterations are related to non-motor symptoms in multiple sclerosis. J. Neurol. **258**(2), 195–202 (2011)

[71] Hsu, P.W., Prager, R., Gee, A., Treece, G.: Freehand 3d ultrasound calibration: A review. In: C. Sensen, B. Hallgrímsson (eds.) Advanced Imaging in Biology and Medicine, pp. 47–84. Springer Berlin Heidelberg (2009)

[72] Hughes, A.J., Daniel, S.E., Kilford, L., Lees, A.J.: Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J. Neurol. Neurosurg. Psychiatr. **55**(3), 181–184 (1992)

[73] Iglesias, J.E., Konukoglu, E., Montillo, A., Tu, Z., Criminisi, A.: Combining generative and discriminative models for semantic segmentation of ct scans via active learning. In: IPMI (2011)

[74] Ionasec, R.I., Voigt, I., Georgescu, B., Wang, Y., Houle, H., Vega-Higuera, F., Navab, N., Comaniciu, D.: Patient-specific modeling and quantification of the aortic and mitral valves from 4-D cardiac CT and TEE. IEEE Trans Med Imaging **29**(9), 1636–1651 (2010)

[75] Ivancevich, N., Dahl, J., Trahey, G., Smith, S.: Phase-aberration correction with a 3-d ultrasound scanner: feasibility study. Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on **53**(8), 1432 –1439 (2006). DOI 10.1109/TUFFC.2006.1665100

[76] Jalali, B.: Multi-modal image registration for pre-operative surgical planning and intra-operative guidance using adaptive distance measures. Master's thesis, Chalmers University of Technology, Department of Signals and Systems (2011)

[77] Jankovic, J.: Parkinson's disease: clinical features and diagnosis. J. Neurol. Neurosurg. Psychiatr. **79(4)**, 368–376 (2008)

[78] Jensen, J.A.: A model for the propagation and scattering of ultrasound in tissue. J. Acoust. Soc. Am. **89(1)**, pp. 182–190 (1991). DOI 10.1121/1.400497

[79] Jensen, J.A.: Field: A program for simulating ultrasound systems. In: 10TH NORDICBALTIC CONFERENCE ON BIOMEDICAL IMAGING, VOL. 4, SUPPLEMENT 1, PART 1:351–353, pp. 351–353 (1996)

[80] Jensen, J.A.: Linear description of ultrasound imaging systems. Notes for the International Summer School on Advanced Ultrasound Imaging, Technical University of Denmark (1999)

[81] JM Bronstein M Tagliati, R.A.: Deep brain stimulation for parkinson disease: an expert consensus and review of key issues. Arch. Neurol. **68(2)**, 165 (2011)

[82] John, J.P., Wang, L., Moffitt, A.J., Singh, H.K., Gado, M.H., Csernansky, J.G.: Inter-rater reliability of manual segmentation of the superior, inferior and middle frontal gyri. Psychiatry Res **148**(2-3), 151–163 (2006)

[83] Johnson, D.H., Dudgeon, D.E.: Array Signal Processing - Concepts and Techniques, chap. Chapter 4 - Beamforming, p. 111 ff. Simon & Schuster (1993)

[84] Karamalis, A., Wein, W., Navab, N.: Fast ultrasound image simulation using the westervelt equation. In: Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention: Part I, MICCAI'10, pp. 243–250. Springer-Verlag, Berlin, Heidelberg (2010)

[85] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. INTERNATIONAL JOURNAL OF COMPUTER VISION **1**(4), 321–331 (1988)

[86] Kaur, J., Kaur, J., Kaur, M.: Survey of Despeckling Techniques for Medical Ultrasound Images. Int. J. Comp. Tech. Appl. **2(4)**, 1003–1007 (2011)

[87] Kier, C., Cyrus, C., Seidel, G., Hofmann, U.G., Aach, T.: Segmenting the substantia nigra in ultrasound images for early diagnosis of parkinson's disease. International Journal of Computer Assisted Radiology and Surgery **2**(S1), 83–85 (2007)

[88] Kirschner, M., Wesarg, S.: Regularisierung lokaler deformation im probabilistischen active shape model. In: T. Tolxdorff, T.M. Deserno, H. Handels, H.P. Meinzer (eds.) Bildverarbeitung für die Medizin 2012, Informatik aktuell, pp. 328–333. Springer Berlin Heidelberg (2012)

[89] Klein, T.J.: Statistical Image Processing of Medical Ultrasound Radio Frequency Data. Ph.D. thesis, Technische Universität München, Chair for Computer Aided Medical Procedures (CAMP), Institut für Informatik I-16, Boltzmannstr. 3, 85748 Garching b. München (2012)

[90] Klotzsch, C., Bozzato, A., Lammers, G., Mull, M., Lennartz, B., Noth, J.: Three-dimensional transcranial color-coded sonography of cerebral aneurysms. Stroke **30**(11), 2285–2290 (1999)

[91] Krissian, K., Kikinis, R., Westin, C.F., Vosburgh, K.: Speckle-constrained filtering of ultrasound images. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02, CVPR '05, pp. 547–552. IEEE Computer Society (2005)

[92] Krissian, K., Westin, C.F., Kikinis, R., Vosburgh, K.: Oriented speckle reducing anisotropic diffusion. Image Processing, IEEE Transactions on **16**(5), 1412 –1424 (2007)

[93] Krucker, J., LeCarpentier, G., Fowlkes, J., Carson, P.: Rapid elastic image registration for 3D ultrasound. IEEE Trans. Med. Imag. **21**(11), 1384–1394 (2002)

[94] Kutter, O., Shams, R., Navab, N.: Visualization and gpu-accelerated simulation of medical ultrasound from ct images. Comput. Methods Prog. Biomed. **94**(3), 250–266 (2009)

[95] Lankton, S., Tannenbaum, A.: Localizing region-based active contours. Image Processing, IEEE Transactions on **17**(11), 2029 –2039 (2008)

[96] Lankton, S., Tannenbaum, A.: Localizing region-based active contours. Image Processing, IEEE Transactions on **17**(11), 2029 –2039 (2008). DOI 10.1109/TIP.2008.2004611

[97] Lees, A.J., Parkinson, J.: Unresolved issues relating to the shaking palsy on the celebration of James Parkinson's 250th birthday. Mov. Disord. **22 Suppl 17**, S327–334 (2007)

[98] Lesage, S., Brice, A.: Parkinson's disease: from monogenic forms to genetic susceptibility factors. Hum. Mol. Genet. **18**(R1), 48–59 (2009)

[99] van de Loo, S., Walter, U., Behnke, S., Hagenah, J., Lorenz, M., Sitzer, M., Hilker, R., Berg, D.: Reproducibility and diagnostic accuracy of substantia nigra sonography for the diagnosis of parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry **81**(10), 1087–1092 (2010)

[100] Lozano, A.M., Giacobbe, P., Hamani, C., Rizvi, S.J., Kennedy, S.H., Kolivakis, T.T., Debonnel, G., Sadikot, A.F., Lam, R.W., Howard, A.K., Ilcewicz-Klimek, M., Honey, C.R., Mayberg, H.S.: A multicenter pilot study of subcallosal cingulate area deep brain stimulation for treatment-resistant depression. J. Neurosurg. **116**(2), 315–322 (2012)

[101] Marczak, W.: Water as a standard in the measurements of speed of sound in liquids. The Journal of the Acoustical Society of America **102(5)**, pp.2776–2779 (1997)

[102] McDonald, W.M., Richard, I.H., DeLong, M.R.: Prevalence, etiology, and treatment of depression in Parkinson's disease. Biol. Psychiatry **54**(3), 363–375 (2003)

[103] Mehnert, S., Reuter, I., Schepp, K., Maaser, P., Stolz, E., Kaps, M.: Transcranial sonography for diagnosis of Parkinson's disease. BMC Neurol **10**, 9 (2010)

[104] Mercier, L., Lang?, T., Lindseth, F., Collins, D.L.: A review of calibration techniques for freehand 3-D ultrasound systems. Ultrasound Med Biol **31**(4), 449–471 (2005)

[105] Michaeli, S., Oz, G., Sorce, D., Garwood, M., Ugurbil, K., Majestic, S.: Assessment of brain iron and neuronal integrity in patients with Parkinson's disease using novel MRI contrasts. Movement disorders : official journal of the Movement Disorder Society **22**(3), 334–340 (2007)

[106] Misaridis, T.: Ultrasound imaging using coded signals. Ph.D. thesis, Center for Fast Ultrasound Imaging, Technical University of Denmark (2001)

[107] Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled decision forests and their application for semantic segmentation of ct images. In: IPMI (2011)

[108] Naumann, M., Becker, G., Toyka, K.V., Supprian, T., Reiners, K.: Lenticular nucleus lesion in idiopathic dystonia detected by transcranial sonography. Neurology **47**(5), 1284–1290 (1996)

[109] Noble, J.A.: Ultrasound image segmentation and tissue characterization. Proc Inst Mech Eng H **224**(2), 307–316 (2010)

[110] Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: a survey. IEEE Trans Med Imaging **25**(8), 987–1010 (2006)

[111] Obeso, J.A., Rodriguez-Oroz, M.C., Goetz, C.G., Marin, C., Kordower, J.H., Rodriguez, M., Hirsch, E.C., Farrer, M., Schapira, A.H., Halliday, G.: Missing pieces in the Parkinson's disease puzzle. Nat. Med. **16**(6), 653–661 (2010)

[112] Parkinson, J., Parkinson, J.: An essay on the shaking palsy. 1817. J Neuropsychiatry Clin Neurosci **14**(2), 223–236 (2002)

[113] Pauly, O., Ahmadi, S.A., Plate, A., Bötzel, K., Navab, N.: Detection of Substantia Nigra Echogenicities in 3D Transcranial Ultrasound for Early Diagnosis of Parkinson Disease. In: Proceedings Internation Conference Medical Image Computing and Computer Aided Intervention (MICCAI), p. (to appear) (2012)

[114] Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Martinez-Moeller, A., Nekolla, S., Navab, N.: Fast multiple organ detection and localization in whole-body mr dixon sequences. In: Proc. MICCAI (2011)

[115] Piella, G., De Craene, M., Yao, C., Penney, G.P., Frangi, A.F.: Multiview diffeomorphic registration for motion and strain estimation from 3DUS sequences. In: Int'l Conf. Funct. Imag. and Modeling of the heart, FIMH, pp. 375–383 (2011)

[116] Pinton, G., Dahl, J., Rosenzweig, S., Trahey, G.: A heterogeneous nonlinear attenuating full- wave model of ultrasound. Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on **56**(3), 474 –488 (2009). DOI 10.1109/TUFFC.2009.1066

[117] Plate, A., Ahmadi, S.A., Klein, T., Navab, N., Weisse, J., Mehrkens, J., Bötzel, K.: Towards a More Objective Visualization of the Midbrain and its Surroundings Using 3D Transcranial Ultrasound. In: 54. Jahrestagung der Deutschen Gesellschaft für Klinische Neurophysiologie und Funktionelle Bildgebung (2010)

[118] Plate, A., Ahmadi, S.A., Klein, T., Pauly, O., Navab, N., Bötzel, K.: A pilot study using 3D-transcranial sonography for (semi-)automated diagnosis of Parkinson´s Disease. In: 84. Jahreskongress der Deutschen Gesellschaft für Neurologie (DGN) (2011)

[119] Plate, A., Ahmadi, S.A., Klein, T., Pauly, O., Navab, N., Bötzel, K.: An objective 3D-transcranial sonography method for detecting parkinson's disease – a pilot study. In: 7. Deutscher Parkinson Kongress. Kiel (2011)

[120] Plate, A., Ahmadi, S.A., Klein, T., Pauly, O., Navab, N., Bötzel, K.: Development of a 3D-transcranial sonography method for detecting Parkinson's disease (PD) – A pilot study. In: 15th International Congress of Parkinson's Disease and Movement Disorders. Movement Disorder Society (MDS) (2011)

[121] Plate, A., Ahmadi, S.A., Pauly, O., Klein, T., Navab, N., Bötzel, K.: 3d sonographic examination of the midbrain for computer-aided diagnosis of movement disorders. Ultrasound in Medicine & Biology (UMB) **(to appear)**, pp. (2012)

[122] Poon, T., Rohling, R.: Three-dimensional extended field-of-view ultrasound. Ultrasound Med. Biol. **32**(3), 357–369 (2006)

[123] Prager, R.W., Ijaz, U.Z., Gee, A.H., Treece, G.M.: Three-dimensional ultrasound imaging. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine **224**(2), 193–223 (2010). DOI 10.1243/09544119JEIM586. URL http://pih.sagepub.com/content/224/2/193.abstract

[124] Prestel, J., Schweitzer, K.J., Hofer, A., Gasser, T., Berg, D.: Predictive value of transcranial sonography in the diagnosis of Parkinson's disease. Mov. Disord. **21**(10), 1763–1765 (2006)

[125] Puls, I., Berg, D., Maurer, M., Schliesser, M., Hetzel, G., Becker, G.: Transcranial sonography of the brain parenchyma: comparison of B-mode imaging and tissue harmonic imaging. Ultrasound Med Biol **26**(2), 189–194 (2000)

[126] R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2008). URL http://www.R-project.org. ISBN 3-900051-07-0

[127] Rangarajan, A., Chui, H., Bookstein, F.L.: The softassign procrustes matching algorithm. In: Proceedings of the 15th International Conference on Information Processing in Medical Imaging,

IPMI '97, pp. 29–42. Springer-Verlag, London, UK, UK (1997). URL http://dl.acm.org/citation.cfm?id=645595.660411

[128] Rasmussen, I.A., Lindseth, F., Rygh, O.M., Berntsen, E.M., Selbekk, T., Xu, J., Nagelhus Hernes, T.A., Harg, E., Haberg, A., Unsgaard, G.: Functional neuronavigation combined with intra-operative 3D ultrasound: initial experiences during surgical resections close to eloquent brain areas and future directions in automatic brain shift compensation of preoperative data. Acta Neurochir (Wien) **149**(4), 365–378 (2007)

[129] Reinertsen, I., Descoteaux, M., Siddiqi, K., Collins, D.L.: Validation of vessel-based registration for correction of brain shift. Med Image Anal **11**(4), 374–388 (2007)

[130] Revelle, W.: psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois (2012). URL http://personality-project.org/r/psych.manual.pdf. R package version 1.2.1

[131] Rodgers, J.L., Nicewander, W.A.: Thirteen ways to look at the correlation coefficient. The American Statistician **42**(1), pp. 59–66 (1988). URL http://www.jstor.org/stable/2685263

[132] Sakalauskas, A., Lukosevicius, A., Lauckaite, K.: Texture analysis of transcranial sonographic images for parkinson disease diagnostics. ULTRAGARSAS (ULTRASOUND), **Vol. 66 (3)**, 32–36 (2011)

[133] Samii, A., Nutt, J.G., Ransom, B.R.: Parkinson's disease. Lancet **363**(9423), 1783–1793 (2004)

[134] Sankar, T., Tierney, T.S., Hamani, C.: Novel applications of deep brain stimulation. Surg Neurol Int **3**(Suppl 1), S26–S33 (2012)

[135] Schapira, A.H., Obeso, J.: Timing of treatment initiation in Parkinson's disease: a need for reappraisal? Ann. Neurol. **59**(3), 559–562 (2006)

[136] Shams, R., Hartley, R., Navab, N.: Real-time simulation of medical ultrasound from ct images. In: Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, Part II, MICCAI '08, pp. 734–741. Springer-Verlag, Berlin, Heidelberg (2008)

[137] Sheffer, A., Praun, E., Rose, K.: Mesh parameterization methods and their applications. Found. Trends. Comput. Graph. Vis. **2**(2), 105–171 (2006)

[138] Shen, L., Farid, H., McPeek, M.A.: Modeling three-dimensional morphological structures using spherical harmonics. Evolution: International Journal of Organic Evolution **63**(4), 1003–1016 (2009)

[139] Shen, L., Farid, H., McPeek, M.A.: Modeling three-dimensional morphological structures using spherical harmonics. Evolution: International Journal of Organic Evolution **63**(4), 1003–1016 (2009)

[140] Shen, L., Firpi, H.A., Saykin, A.J., West, J.D.: Parametric surface modeling and registration for comparison of manual and automated segmentation of the hippocampus. Hippocampus **19**(6), 588–595 (2009)

[141] Shen, L., Huang, H., Makedon, F., Saykin, A.J.: Efficient registration of 3d spharm surfaces. In: Proceedings of the Fourth Canadian Conference on Computer and Robot Vision, CRV '07, pp. 81–88. IEEE Computer Society, Washington, DC, USA (2007)

[142] Shen, L., Makedon, F.: Spherical mapping for processing of 3d closed surfaces. Image and Vision Computing **24(7)**, 743–761 (2006)

[143] Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference, ACM '68, pp. 517–524. ACM, New York, NY, USA (1968)

[144] Shoemake, K.: Animating rotation with quaternion curves. SIGGRAPH Comput. Graph. **19**(3), 245–254 (1985)

[145] Shrout, P.E., Fleiss, J.L.: Intraclass correlations: uses in assessing rater reliability. Psychol Bull **86**(2), 420–428 (1979)

[146] Sielhorst, T., Feuerstein, M., Traub, J., Kutter, O., Navab, N.: Campar: A software framework guaranteeing quality for medical augmented reality. Int. J. Comput. Assist. Radiol. Surg **1**, 29–30 (2006)

[147] Singer, C.: Managing the patient with newly diagnosed Parkinson disease. Cleve Clin J Med **79 Suppl 2**, 3–7 (2012)

[148] Sivakumar, R., Nedumaran, D.: Comparative study of Speckle Noise Reduction of Ultrasound B-scan Images in Matrix Laboratory Environment. Internation Journal of Computer Applications **10(9)**, 46–50 (2010)

[149] Skoloudik, D., Fadrna, T., Bartova, P., Langova, K., Ressner, P., Zapletalova, O., Hlustik, P., Herzig, R., Kannovsky, P.: Reproducibility of sonographic measurement of the substantia nigra. Ultrasound Med Biol **33**(9), 1347–1352 (2007)

[150] Skoloudik, D., Walter, U.: Method and validity of transcranial sonography in movement disorders. Int. Rev. Neurobiol. **90**, 7–34 (2010)

[151] Slabaugh, G., Unal, G.: Active polyhedron: surface evolution theory applied to deformable meshes. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2, pp. 84 – 91 vol. 2 (2005). DOI 10.1109/CVPR.2005.60

[152] Tang, S.C., Clement, G.T., Hynynen, K.: A computer-controlled ultrasound pulser-receiver system for transskull fluid detection using a shear wave transmission technique. IEEE Trans Ultrason Ferroelectr Freq Control **54**(9), 1772–1783 (2007)

[153] Tay, P., Garson, C., Acton, S., Hossack, J.: Ultrasound despeckling for contrast enhancement. Image Processing, IEEE Transactions on **19**(7), 1847 –1860 (2010). DOI 10.1109/TIP.2010.2044962

[154] Thodberg, H.H.: Minimum description length shape and appearance models. In: In Image Processing Medical Imaging, IPMI, pp. 51–62. Springer (2003)

[155] Tong, S., Cardinal, H.N., McLoughlin, R.F., Downey, D.B., Fenster, A.: Intra- and inter-observer variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging. Ultrasound Med Biol **24**(5), 673–681 (1998)

[156] Unsgaard, G., Rygh, O.M., Selbekk, T., Muller, T.B., Kolstad, F., Lindseth, F., Hernes, T.A.: Intra-operative 3D ultrasound in neurosurgery. Acta Neurochir (Wien) **148**(3), 235–253 (2006)

[157] various: Parkinson's Disease, chap. Symptomatic pharmacological therapy in Parkinson's disease, pp. pp. 59–100. The National Collaborating Centre for Chronic Conditions, Royal College of Physicians, London (2006)

[158] various: Parkinson's Disease, chap. Surgery for Parkinson's disease, pp. pp. 59–100. The National Collaborating Centre for Chronic Conditions, Royal College of Physicians, London (2006)

[159] Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vision **57**(2), 137–154 (2004). DOI 10.1023/B:VISI.0000013087.49260.fb. URL http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb

[160] Vlaar, A., Hovestadt, A., van Laar, T., Bloem, B.R.: The treatment of early Parkinson's disease: levodopa rehabilitated. Pract Neurol **11**(3), 145–152 (2011)

[161] Vlaar, A., de Nijs, T., van Kroonenburgh, M., Mess, W., Winogrodzka, A., Tromp, S., Weber, W.: The predictive value of transcranial duplex sonography for the clinical diagnosis in undiagnosed parkinsonian syndromes: comparison with spect scans. BMC Neurology **8**(1), 42 (2008). DOI 10.1186/1471-2377-8-42

[162] Vlaar, A., Tromp, S.C., Weber, W.E., Hustinx, R.M., Mess, W.H.: The reliability of transcranial duplex scanning in parkinsonian patients: comparison of different observers and ultrasound systems. Ultraschall Med **32 Suppl 1**, S83–88 (2011)

[163] Vlaar, A.M., Bouwmans, A., Mess, W.H., Tromp, S.C., Weber, W.E.: Transcranial duplex in the differential diagnosis of parkinsonian syndromes: a systematic review. J. Neurol. **256**(4), 530–538 (2009)

[164] Wachinger, C., Klein, T., Navab, N.: The 2D analytic signal for envelope detection and feature extraction on ultrasound images. Med Image Anal **(to appear)**, pp. (2012)

[165] Wachinger, C., Wein, W., Navab, N.: Three-dimensional ultrasound mosaicing. In: MICCAI, pp. 327–335 (2007)

[166] Waelkens, P., Ahmadi, S.A., Navab, N.: 'frangi goes us': Multiscale tubular structure detection adapted to 3d ultrasound. In: Proc. International Conference on Medical Image Computing and Computer Assisted Intervention 2012 (MICCAI) (to appear) (2012)

[167] Walter, U., Behnke, S., Eyding, J., Niehaus, L., Postert, T., Seidel, G., Berg, D.: Transcranial brain parenchyma sonography in movement disorders: state of the art. Ultrasound Med Biol **33**(1), 15–25 (2007)

[168] Walter, U., Dressler, D., Probst, T., Wolters, A., Abu-Mugheisib, M., Wittstock, M., Benecke, R.: Transcranial brain sonography findings in discriminating between parkinsonism and idiopathic parkinson disease. Archives of Neurology **64**(11), 1635–1640 (2007)

[169] Walter, U., Dressler, D., Wolters, A., Probst, T., Grossmann, A., Benecke, R.: Sonographic discrimination of corticobasal degeneration vs progressive supranuclear palsy. Neurology **63**(3), 504–509 (2004)

[170] Walter, U., Niehaus, L., Probst, T., Benecke, R., Meyer, B.U., Dressler, D.: Brain parenchyma sonography discriminates Parkinson's disease and atypical parkinsonian syndromes. Neurology **60**(1), 74–77 (2003)

[171] Walter, U., Wagner, S., Horowski, S., Benecke, R., Zettl, U.K.: Transcranial brain sonography findings predict disease progression in multiple sclerosis. Neurology **73**(13), 1010–1017 (2009)

[172] Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging **23**(7), 903–921 (2004)

[173] Wein, W.: Multimodal Integration of Medical Ultrasound for Treatment Planning and Interventions. Ph.D. thesis, Technische Universität München, Fakultät für Informatik (2007)

[174] Wein, W., Brunke, S., Khamene, A., Callstrom, M., Navab, N.: Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis **12**(5), 577–585 (2008)

[175] Wein, W., Pache, F., Roeper, B., Navab, N.: Backward-Warping Ultrasound Reconstruction for Improving Diagnostic Value and Registration. In: MICCAI 2006 Proceedings, Lecture Notes in Computer Science. Springer (2006)

[176] White, P.J., Whalen, S., Tang, S.C., Clement, G.T., Jolesz, F., Golby, A.J.: An intraoperative brain shift monitor using shear mode transcranial ultrasound: preliminary results. J Ultrasound Med **28**(2), 191–203 (2009)

[177] Woo, J., Hong, B.W., Hu, C.H., Shung, K.K., Kuo, C.C., Slomka, P.J.: Non-rigid ultrasound image registration based on intensity and local phase information. J. Signal Process. Syst. **54**(1-3), 33–43 (2009)

[178] Yaqub, M., Ioannou, C., Papageorghiou, A., Javaid, K., Cooper, C., Noble., J.: Improving boundary definition for 3D ultrasound quantification of fetal femur. In: Proc. MIAU, pp. 89–93 (2010)

[179] Yu, H., Pattichis, M., Goens, M.: Robust segmentation and volumetric registration in a multi-view 3D freehand ultrasound reconstruction system. In: Signals, Systems and Computers ACSSC, pp. 1978 –1982 (2006)

[180] Yu, Y., Acton, S.: Speckle reducing anisotropic diffusion. Image Processing, IEEE Transactions on **11**(11), 1260 – 1270 (2002)

[181] Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage **31**(3), 1116–1128 (2006)

[182] Zhu, Y., Magee, D., Ratnalingam, R., Kessel, D.: A training system for ultrasound-guided needle insertion procedures. In: Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention (MICCAI) - Volume Part I, MICCAI'07, pp. 566–574. Springer-Verlag, Berlin, Heidelberg (2007)

[183] Zikic, D., Wein, W., Khamene, A., Clevert, D.A., Navab, N.: Fast deformable registration of 3D-ultrasound data using a variational approach. In: MICCAI, pp. 915–923 (2006)

[184] Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Kikinis, R.: Statistical validation of image segmentation quality based on a spatial overlap index. Acad Radiol **11**(2), 178–189 (2004)

[185] Zweig, M.H., Campbell, G.: Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. **39**(4), 561–577 (1993)