Improving image-guided preoperative planning in deep brain stimulation procedures

By

Yuan Liu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

December, 2016

Nashville, Tennessee

Approved by:

Professor Benoit M. Dawant
Professor Bennett A. Landman
Professor Jack H. Noble
Professor Michael I. Miga
Professor Peter E. Konrad

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and support from my committee members and the assistance from my colleagues at Vanderbilt.

First and foremost, I owe the deepest gratitude to my advisor, Dr. Benoit Dawant. He introduced me to this amazing research area and gave me the opportunity to be part of the team with financial support throughout the years. He equipped me with advanced image processing techniques and gave me room to explore research topics that I am interested in. His guidance has helped me progress both professionally and personally. His honest, hardworking, and attention-to-detail work ethic and positive attitude are true inspirations from which I will benefit for life.

I would also like to express my appreciation to other committee members. Dr. Jack Noble shared his image processing expertise and gave valuable suggestions to my work, especially in my early days at Vanderbilt. His tool IMPROVISER is extremely useful for manual delineation and visualization of medical image data. Dr. Konrad, despite his busy schedule as a neurosurgeon, always makes the time to conduct clinical experiments and provides feedback to my work. Through those meetings, I have learned a lot of anatomical knowledge from him. Thanks to Dr. Bennett Landman for imparting the knowledge. I took three courses with him, and those have expanded my horizon and inspire me for research ideas. I would also like to thank Dr. Michael Miga for his feedback on my work.

Thank for my colleagues in the DBS team. I want to thank Dr. Pierre-Fran ̧ois D'Haese in particular. He is a main contributor to many of my research projects, especially my last and latest, the thalamic nuclei segmentation project. Without his extensive knowledge of neuroanatomy and his immense effort in delineating those fine contours, this project will not be

possible. Thanks for Dr. Srivatsan Pallavaram, Rui Li, William J. Rodriguez, and Priyanka Prasad, for helping me understand the complex DBS system and the internal tools, which I found very efficient to use and easy to navigate. Thanks for the neurosurgeons Dr. Neimat, Dr. Yu, and Dr. Tatter, for their valuable time and input to those clinical experiments.

I also want to thank the former lab members, Zhaoying Han, Antong Chen, Siyi Ding, Ryan Datteri, Ankur Kumar, and Fitsum Reda, for their stimulating discussions and sincere friendship. Thanks for my current lab members; you have made my experience enjoyable.

Special thanks to financial support by NIH grants R01-EB006136, R01-NS095291, and Fellowship from Vanderbilt Institute in Surgery and Engineering.

Last but not the least, I would like to thank my parents, for their infinite love and support throughout my life, and my boyfriend, who understands me the most and has shared all the moments of sweetness and bitterness. I dedicate my work to all of you.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1. Background on Deep Brain Stimulation

Parkinson's disease (PD), essential tremor (ET), dystonia, and obsessive-compulsive disorder (OCD) are chronic neurological disorders that affect as many as 1 million, 10 million, 250,000, and 2.2 million people respectively in the United States [1-4]. Deep brain stimulation (DBS), a stereotactic functional neurosurgery first introduced in 1987, has proven effective in symptom relief for those otherwise-treatment-resistant movement and mental disorders. Since the approval by the Food and Drug Administration (FDA) for ET in 1997, for PD in 2002, for dystonia in 2003, and for OCD in 2009, it has been used to treat an increasing number of patients. In this procedure, electrodes are implanted in the deep brain to stimulate target brain nucleus with high frequency electrical pulses. Although the underlying mechanism is not completely understood, the goal of this intervention is to interfere with surrounding fiber pathways and modulate the brain network responsible for these diseases [5]. Unlike those of ablation techniques, its effects are adjustable and reversible without any damage to healthy brain tissue.

There are several stages involved in DBS procedure, i.e., preoperative planning, intraoperative placement, and postoperative programming [6]. In the preoperative stage, neurosurgeons carefully examine the patient's anatomical images and use their experience and knowledge to make a plan for the intervention. This plan mainly includes the identification of optimal a) target position for stimulation and b) linear trajectory from the outer surface of the skull to the chosen target for implantation. During the surgery, a small burr hole is drilled at the

planned entry point on the skull to have access to the target region. A test electrode is inserted at first along the planned trajectory and probed around the planned target with varying depths at micrometric precision. During this probing process, micro-electrode recordings (MER) of neuronal activity and micro/macro- stimulation responses from the awake patient are collected. These complementary measurements allow the whole surgical team (neurosurgeons, neurologists, and neurophysiologists) to establish the functional borders of the deep brain nuclei and finalize the target location that has the best tradeoff between therapeutic benefits and side effects [7-8]. Then a permanent four-contact electrode is implanted at the refined target position. As shown in Figure 1.1, this electrode is connected via extension wire to a battery-powered pulse generator placed in a subclavicular subcutaneous pocket. Finally, in the postoperative programming stage, the optimal contact(s) of the electrode is(are) identified and the pulse generator is calibrated with optimal voltage or current configurations to maximize symptom suppression and control side effects over several weeks after the surgery.

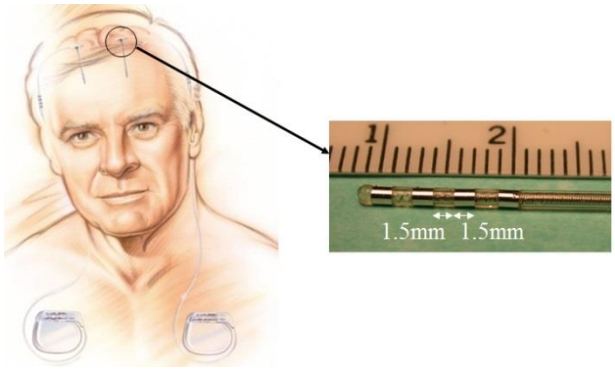

Figure 1.1. Illustration of the DBS system. The pulse generator is connected through extension wire to the lead implanted in the deep brain. An example of the electrode is shown on the right (Medtronic #3387 quadripolar lead).

DBS is an efficient but complex therapy with little tolerance for error. Accurate electrode placement at a millimetric precision level has been found to contribute to improved clinical

outcomes and reduced occurrence of adverse effects [9-10]. In current practice, this is achieved by preoperatively selecting the target and the trajectory based on visual inspection of anatomic images and intraoperatively refining the target using MER and stimulation responses. Without proper planning in advance, the intervention can be prolonged for hours with the patient awake during the trial-and-error search for the optimal target. Moreover, risks of surgical complications related to lead insertions may be increased, which could cause potential brain tissue damage or internal bleeding. Consequently, preoperative planning of a safe insertion trajectory that reaches a precise target is essential for the success of the DBS procedure.

In the preoperative planning phase, a single or multiple set of magnetic resonance (MR) volumes and a computed tomography (CT) scan are typically acquired for each patient. Manual planning with such datasets is a complex, laborious, and subjective process. A number of neuronagivation platforms have been developed to assist neurosurgical teams in these tasks, which include CranialVault [11], CAPS [12], and PyDBS [13]. These systems allow users to easily navigate through co-registered image slices and offer additional image computing functionalities to ease planning tasks. Automatically finding optimal trajectories is difficult because of meticulous operating requirements for safe electrode insertion and limitations in clinical imaging qualities. In the past few decades, numerous research efforts have been made to provide interactive or automatic solutions, and furthermore, conduct retrospective and/or perspective clinical validation of those solutions.

In the following sections, we will present background information and brief reviews on DBS preoperative planning. This is separated into two tasks. Given a pre-identified target point, we will first focus on finding the direction of the electrode that forms an optimal linear trajectory, which we refer to as trajectory planning. We will then focus on various approaches for finding an

accurate target point, which we refer to as target localization. Finally, we will present the goal and the contribution of the dissertation.

1.2. Trajectory Planning

An ideal trajectory should satisfy a number of surgical requirements, such as avoiding certain critical structures and staying within certain angle range confined by the stereotactic frame. Manual selection of this trajectory involves a neurosurgeon empirically picking a few candidates and examining each of them by slicing through the preoperative CT/MRI head scans and alternating different sequences from time to time. This requires the surgeon to mentally identify and reconstruct complex 3D structures and derive spatial relationships between surrounding structures and the trajectory. Computer-assistance is highly appreciated to facilitate this process.

Over the years, various trajectory planning methods have been proposed for DBS and other keyhole neurosurgical procedures to mimic surgeon's decision-making process [14-20]. Early efforts by Vaillant *et al.* [14] and Navkar *et al.* [16] focus on developing interactive tools to visualize critical structures and their proximity to manual selections. In addition to that, recent works aim to provide analytical solutions by automatically ranking a large number of candidate trajectories according to some surgical constraints and their relative importance [18-20]. These solutions vary in terms of the choice of constraints, the algorithms used to segment relevant structures, and/or the mathematical formulation of those constraints. Common constraints consist of optimizing the distances to vessels, sulci, and ventricles, while other rules include restrictions in path length, lead orientation, and overlap with the caudate [19-20]. A variety of segmentation methods are applied to delineate those critical structures, including single or multi-atlas-based segmentations or other approaches that exploit multi-modality image information of target

subjects. To quantify the risk for each structure, costs for all voxels along the trajectory are derived from the segmented structure's distance map and are summarized by taking either the sum [14], the maximum [15][19], or both [18][20]. Risks for different constraints are then weighted based on their surgical importance. A linear aggregation of those risks is typically used to produce a final score for each trajectory.

The absence of a ground truth poses challenges to assess whether those automatic methods are superior to manual protocol. Retrospective and prospective evaluation studies have been conducted both qualitatively and quantitatively [12][14-20]. Qualitative validation involves user-experience questionnaires about effectiveness of the interactive software [15-16] and comparative experiments in which automatic trajectories are rated against manual ones by surgeons [20]. Quantitative analysis of the automatic trajectories is also performed by listing their distances to critical structures [18-20] and the costs for each individual constraint [19-20]. However, such analysis assumes that the surgical constraints are perfectly formulated to describe exactly what surgeons seek for. Moreover, segmentation errors need to be ruled out for the quantitative measurements presented above to be reliable.

Importantly, current trajectory planning frameworks are designed and evaluated by the same surgeons. The surgical constraints are defined based on rules expressed by surgeons during interviews, and the associated weights are usually set heuristically based on their preferences. Further investigations are necessary to check if adjustments in those constraints or weights need to be made for the same system to be used by different surgeons at different institutes. The focus of our work presented herein is to not only develop an automatic trajectory planning system, but also evaluate the clinical effectiveness of this system in a multi-surgeon, multi-site setting.

1.3. Target Localization

As mentioned earlier, accurate initialization of the surgical target may reduce the need for invasive intraoperative exploration, and thereby decrease the surgical duration and procedure-related complications. Common target nuclei of interest include the thalamic ventral intermediate nucleus (VIM) [21], subthalamic nucleus (STN) [22], and globus pallidus internus (GPi) [22] for treating different symptoms. Segmentation of these structures is valuable to guide the surgeons in target selection, but is challenging to obtain because of their small size and poor contrast in clinical CT or T1-weighted (T1-w) images. T2-weighted (T2-w) sequences have sometimes been used to visualize the STN, but its insufficient and inconsistent contrast across patients makes it a less reliable choice. Moreover, there is a lack of consensus on the exact location within these structures for optimal stimulation. Over the years, a number of approaches have been proposed to either segment those structures or predict the position of the target point. Those are aided by complementary information from visible landmarks, conventional brain atlases, functional data from past patients, and advanced imaging sequences. In the following subsections we will review different types of state-of-the-art targeting approaches based on their operating mechanisms and introduce our contributions for each subcategory.

1.3.1. AC-PC-based Targeting

Normal clinical practice relies on visible landmarks in T1-w images to indirectly determine the target position. This is typically done by taking their standardized coordinates in a stereotactic system defined by the Anterior Commissure (AC), Posterior Commissure (PC), and midsagittal plane (MSP). Establishing this spatial reference system not only enables the identification of the targets, but also permits effective communication of their locations in neurosurgical community.

This also serves as a crucial step for many medical imaging tasks such as registration initialization and human brain mapping [24-25].

Inter-surgeon variability in manually selecting the AC, PC, and MSP can have a substantial effect on DBS targeting [26]. Automatic approaches that accurately and consistently localize the landmarks and plane across patients can aid in standardizing this coordinate system. This is challenging because of image contaminations, inter-subject anatomical variations, and the non-planarity of the MSP. Most existing algorithms for localizing the AC and PC perform initialization based on surrounding structures or landmarks such as the corpus callosum and midbrain-pons junction, which requires them to be successfully identified [27-31]. Edge detection used in Han *et al.* [27] and Verard *et al.* [28] lacks robustness when noise or motion artifact is present in the image. Several works use image registration to transfer the AC/PC position from the atlas to the patient, which may suffer from long runtime [32-33]. To extract the MSP, methods are primarily divided into two groups, i.e., symmetry-based and feature-based approaches. Symmetry-based approaches assume bilateral symmetry of the brain and search for a plane that maximizes a symmetry measure [34-35]. Accuracy of such methods is limited by the degree of brain asymmetry, especially for patients with schizophrenia, epilepsy, and Alzheimer's disease. Feature-based approaches attempt to identify the inter-hemispheric fissure based on its textural features and fit a plane to the fissure lines [36-37]. Those methods in general are more robust to brain asymmetry but more sensitive to outliers in the extracted features.

To accurately, efficiently, and consistently localize the AC, PC, and MSP, we developed an algorithm to learn their positions in T1-w images from a large training set. Specifically, a univariate regression model is built to associate the appearance of each voxel with its absolute distance to those structures. This algorithm was validated thoroughly via a set of robustness tests

and comparison to other state-of-the-art methods.

1.3.2. Atlas-based Targeting

A major drawback for AC-PC-based targeting is the presumption of a fixed coordinate as target position, which lacks consideration of anatomical variation across patients. To overcome this problem, atlas-based targeting approaches create brain atlases that contain target information and spatially deform those atlases to each individual brain for target prediction. Such atlases can be roughly divided into anatomical atlases and statistical atlases.

Traditional anatomical atlases refer to histological atlases in print or digital formats, with or without an underlying MR volume. They contain a series of unevenly spaced brain sections histologically stained to reveal the structures and substructures of interest [23]. A 3D volume is reconstructed from those 2D sections and spatially aligned to the patient's images for targeting assistance. Methods have been developed to optimize the reconstruction of the 3D atlas volume and the registration between the atlas and patient image volumes [38-39]. For example, Chakravarty *et al.* [38] built a 3D anatomical atlas from a set of serial histological data and registered it to patient volumes via pseudo-MR images. Shortcomings of such methods include brain structure discontinuity, limited anatomical variability, and little to none functional information, which restrict their applicability in DBS targeting.

Statistical atlases are developed to complement existing anatomical atlases with statistical representation of the surgical targets from multiple patients. They are created by mapping each patient to a reference brain template, so that individual electrophysiological recordings or actual implanted positions are transferred to this standardized space. By aggregating functional data from numerous previous patients, they can provide probabilistic maps of actual population-based

target information and deep brain electrophysiological activity. A number of such atlases have been proposed, with variations in the functional data, the registration method, and the aggregation strategy being used for the atlas creation [40-42]. As indicated by Guo *et al.* [42], targeting using such atlases is advantageous compared to those using anatomical atlases. A recent validation study has shown that a statistical atlas-based method can provide fully automatic targeting that matches accuracy of best manual strategies [43].

The success of atlas-based targeting hinges on the nonrigid registration algorithm to generate an accurate spatial mapping between the patient and the reference brain template [44]. Nonrigid registration is an extensively explored topic with a multitude of off-the-shelf solutions available today. Evaluating the performance of those solutions is challenging because of the lack of gold-standard, the usual unavailability of ground truth, and their task-dependent performance. For DBS, earlier evaluation studies of nonrigid registration algorithms for anatomical atlas-based targeting approach focused on segmentation of the STN structure [45-46]. It remains unknown whether the use of different registration algorithms will affect the quality of statistical atlases. To answer this question, we compared the centroid, spread, and shape of statistical atlases created according to the description in [43] but with different registration methods in a large-scale study.

1.3.3. Direct Localization Learning-based Targeting

Atlas-based targeting depends on image registration techniques to establish a one-to-one correspondence between structures surrounding the target region in the atlas and those in the patient images. Such techniques typically take a long runtime and require a large parameter set to be adjusted. Supervised learning, on the other hand, may substitute registration approaches by learning this correspondence from a large training dataset. In line with this assumption, we

propose to learn target locations from past patients who underwent DBS surgery. This is different from the learning-based approach we described in section 1.3.1, which indirectly determines target locations by learning the AC, PC, and MSP location.

Direct learning of the target position is complicated by the high degree of intensity homogeneity in the deep brain region. Unlike easily distinguishable landmarks, we cannot rely on discernable features to distinguish the target point from its adjacent neighbors, as we have done for AC/PC/MSP localization. Surrounding anatomy with high contrast, on the other hand, can be informative about the target. For example, AC-PC-based targeting approaches localize the target using its relative distance to the AC, PC, and MSP. Castro *et al.* have shown that segmentations of the lateral ventricle, third ventricle, and interpeduncular cistern could improve targeting accuracy [46]. While these methods rely on specific pre-defined structures in the atlases, it may be beneficial to allow all surrounding structures to vote for the target location and aggregate those votes in a weighted fashion for the final estimate. This can be achieved by multi-variant regression learning, which regresses the appearance of each voxel to a displacement vector of this voxel to the target position. This idea has been used to address similar issues in other applications, such as localizing candidate point in active shape models on 2D radiographs [47] and the center of vertebrae in MR images [48]. To this end, we developed a learning-based system to directly localize the target using a multi-variant regression model. Spatial prior as well as multi-modal contextual information from T1-w and T2-w sequences was incorporated in this system for STN targeting.

1.3.4. Advanced Imaging Sequences-based Targeting

Advanced imaging techniques have been developed to enhance the contrast in the deep brain

region so that direct visualization or segmentation of target nuclei can be achieved. For the STN and GPi, susceptibility-weighted images (SWI) (magnitude, phase, magnitude-phase-fusion images), T2* mapping, DESPOT1 and DESPOT2 sequences haven been proposed as improvement over standard T2-w [49]. For the VIM, visual identification cannot be readily achieved. Instead, parcellation of the thalamus is normally performed to display this internal nucleus. Traynor *et al.* relied on multiple T1 and T2 maps to reveal subtle differences between thalamic nuclei and employed a modified generic algorithm for thalamic clustering [50]. Diffusion weighted imaging (DWI) has been commonly used for this purpose to distinguish voxels inside the thalamus based on local tensor inhomogeneity or structural connectivity to different cortical regions [51-53]. In addition, other approaches exploit the functional connectivity in the thalamo-cortical system, either via task-related function MRI (fMRI) [54] or resting state fMRI [55]. These techniques, however, are limited by the low spatial resolution and signal-to-noise ratio (SNR) of clinical sequences. Their results also need further validation because of the lack of a ground truth.

Recent advances in high-field MR imaging systems has led to improved image quality with finer spatial resolution, higher SNR, and enhanced contrast. This allows small structures which cannot be well delineated at lower field to be visible and segmented much more accurately. For instance, Wang *et al.* demonstrated the use of 7T T1-w and T2-w images for improved visualization of STN and GPi [57]. Newton *et al.* proposed to visualize intrathalamic substructures by combining contrast from a specially designed magnetization-prepared rapid gradient-echo (MPRAGE) sequence and SWI [56] at 7T. Although 7T MR imaging is yet to be approved for clinical application, it permits accurate delineation of target structures and the construction of high quality anatomical atlases from many subjects. This is in contrast to the

histological atlases as mentioned in section 1.3.2 that are limited in quantity.

With a set of high-field atlases, standard atlas-based approaches using single or multiple atlases can be readily applied to segment the target nuclei in clinical images [57]. These techniques are voxel-based solutions that do not account for a-priori shape information, which may not be optimal for segmenting small structures. Alternatively, such prior information can be learned by exploiting the relative position of the target structures and neighboring structures that are visible at clinical lower-field MR images. This has been explored in a number of works, including a recently proposed regression forests-based approach which predicts the STN from a set of pre-segmented structures [58]. To segment thalamic nuclei, we proposed to build statistical shape models using 7T thalamus atlases obtained with a multi-sequence approach and rely on pre-segmented thalamus to infer its internal subdivision. Results were compared to those obtained with single-atlas-based and multi-atlas-based approaches.

1.4. Goals and Contributions of the Dissertation

The goal of this dissertation is to automatically and accurately plan a trajectory and localize the target for the preoperative stage in DBS, which is achieved by a series of methods with thorough validation. The proposed methods form an automatic preoperative planning framework which could be potentially integrated into a clinical workflow. Various targeting approaches that operate on different principles also make the quality control process possible.

Specifically the contributions of this dissertation are summarized below:

Chapter II presents an automatic trajectory planning system which translates a set of surgical constrains into a mathematical model and exhaustively searches for the optimal

trajectory. This system is further validated in a set of experiments with multiple surgeons across different institutions.

Chapter III proposes a learning-based algorithm to automatically and efficiently localize the AC, the PC, and the MSP for indirect targeting. Robustness of this algorithm is tested under varying degrees of brain asymmetry, noise level, and image rotation. Its performance is also compared against traditional atlas-based methods and a publicly available toolkit.

Chapter IV studies the effect of different nonrigid registrations algorithms on the creation of statistical atlases using several reference volumes. Results show no significant differences of those algorithms for target prediction using statistical atlas-based techniques.

Chapter V develops a learning-based technique to directly localize the target positions using spatial prior and multi-modal information. This technique is compared to other targeting methods used or developed in chapter III and IV and shown to achieve a superior or comparable accuracy level.

Chapter VI designs a protocol to create histological-like atlases from 7T high-resolution multi-modal images with manually delineated thalamic nuclei substructures. Single-atlas-based, multi-atlas-based, and statistical shape-based approaches are proposed to use these atlases to segment the internal structures on clinical 3T volumes and evaluated in a leave-one-out validation study.

Chapter VII provides the summary of the work and discusses possible future work.

References

[1] http://www.pdf.org/understanding_pd

[2] http://www.essentialtremor.org/wp-content/uploads/2013/07/FactSheet012013.pdf

[3] http://www.medtronic.com/us-en/patients/conditions/brain-dystonia.html

[4] http://www.adaa.org/about-adaa/press-room/facts-statistics

[5] S. Miocinovic, S. Somayajula, S. Chitnis, and J. L. Vitek, "History, applications, and mechanisms of deep brain stimulation," *JAMA Neurol*, vol. 70, no. 2, pp. 163-171, 2013.

[6] A. L. Benabid, S. Chabardes, J. Mitrofanis, and P. Pollak, "Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease," *Lancet Neurol*, vol. 8, no. 1, pp. 67-81, 2009.

[7] S. J. Groiss, L. Wojtecki, M. Südmeyer, and A. Schnitzler, "Deep brain stimulation in Parkinson's disease," *Ther Adv Neurol Disord*, vol. 2, no. 9, pp. 20-28, 2009.

[8] G. Deuschl, R. Wenzelburger, F. Kopper, and J.Volkmann, "Deep brain stimulation of the subthalamic nucleus for Parkinson's disease: a therapy approaching evidence-based standards, " *J. Neurol*, vol. 250, no. 1, pp. i43-i49, 2003.

[9] M. K. York, E. A. Wilde, R. Simpson, and J. Jankovic, "Relationship between neuropsychological outcome and DBS surgical trajectory and electrode location," *J Neurol Sci*, vol. 287, no. 1, pp. 159–171, 2009.

[10] F. Lalys, C. Haegelen, M. Mehri, S. Drapier, M. Verin, and P. Jannin, "Anatomo-clinical atlases correlate clinical data and electrode contact coordinates: application to subthalamic deep brain stimulation," *J Neurosci Methods*, vol. 212, no. 2, pp. 297–307, 2013.

[11] P.-F. D'Haese, S. Pallavaram, R. Li, M. S. Remple, C. Kao, J. S. Neimat, P. E. Konrad, and B. M. Dawant, "CranialVault and its CRAVE tools: a clinical computer assistance

system for deep brain stimulation therapy," *Med Image Anal*, vol. 16, no. 3, pp. 744–753, 2012.

[12] S. Bériault, S. Drouin, A. F. Sadikot, Y. Xiao, D. L. Collins, and G. B. Pike, "A Prospective Evaluation of Computer-Assisted Deep Brain Stimulation Trajectory Planning," in *Proc. CLIP*, 2013, pp. 42-49.

[13] D. Tiziano, C. Haegelen, C. Essert, S. Fernández-Vidal, F. Lalys, and P. Jannin, "PyDBS: an automated image processing workflow for deep brain stimulation surgery," *Int J Comput Assist Radiol Surg*, vol. 10, no. 2, pp. 117-128, 2015.

[14] M. Vaillant, C. Davatzikos, R. H. Taylor, and R. N. Bryan, "A path-planning algorithm for image-guided neurosurgery," in *Proc. CVRMed-MRCAS*, 1997, pp. 467-476.

[15] E. J. L. Brunenberg, A. Vilanova, V. Visser-Vandewalle, Y. Temel. L. Ackermans, B. Platel, and B. M. Romeny, "Automatic trajectory planning for deep brain stimulation: a feasibility study," in *Proc. MICCAI*, 2007, pp. 584–592.

[16] N. V. Navkar, N. V. Tsekos, J. R. Stafford, J. S. Weinberg, and Z. Deng, "Visualization and planning of neurosurgical interventions with straight access," in *Proc. IPCAI*, 2010, pp. 1-11.

[17] W. L. Nowinski, B. C. Chua, I. Volkau, F. Puspitasari, Y. Marchenko, V. M. Runge, and M. V. Knopp, "Simulation and assessment of cerebrovascular damage in deep brain stimulation using a stereotactic atlas of vasculature and structure derived from multiple 3- and 7-tesla scans: Technical note," *J Neurosurg*, vol. 113, no. 6, pp.1234-1241, 2010.

[18] R. R. Shamir, L. Joskowicz, I. Tamir, E. Dabool, L. Pertman, A. Ben-Ami, and Y. Shoshan, "Reduced risk trajectory planning in image-guided keyhole neurosurgery," *Med Phys*, vol. 39, no. 5, pp. 2885–2895, 2012.

[19] C. Essert, C. Haegelen, F. Lalys, A. Abadie, and P. Jannin, "Automatic computation of electrode trajectories for deep brain stimulation: A hybrid symbolic and numerical approach," *Int J Comput Assist Radiol Surg*, vol. 7, no. 4, pp. 517–532, 2012.

[20] S. Bériault, F. A. Subaie, D. L. Collins, A. F. Sadikot, and G. B. Pike. "A multi-modal approach to computer-assisted deep brain stimulation trajectory planning," *Int J Comput Assist Radiol Surg*, vol. 7, no. 5, pp. 687–704, 2012.

[21] E. Papavassiliou, G. Rau, S. Heath, A. Abosch, N.M. Barbaro, P. S. Larson, K. Lamborn, and P. A. Starr, "Thalamic deep brain stimulation for essential tremor: relation of lead location to outcome, " *Neurosurgery*, vol. 54, p. 1120:1130, 2004.

[22] P. A. Starr, "Placement of deep brain stimulators in subthalamic nucleus or globus pallidus internus: Technical approach," *Stereotact Funct Neurosurg*, vol. 79, no. 3-4, pp. 118-145, 2002.

[23] G. Shaltenbrand, and W. Wahren, *Guide to the Atlas for Stereotaxy of the Human Brain*. Stuttgart, Germany: Thieme, 1977.

[24] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," *J Comput Assist Tomogr*, vol. 18, no. 2, pp. 192-205, 1994.

[25] W. L. Nowinski, and A. Thirunavuukarasuu, "Atlas-assisted localization analysis of functional images," *Med Image Anal*, vol. 5, no. 3, pp. 207-220, 2001.

[26] S. Pallavaram, H. Yu, J. Spooner, P.-F. D'Haese, B. Bodenheimer, P. E. Konrad, and B. M. Dawant, "Intersurgeon variability in the selection of anterior and posterior commissures and its potential effects on target localization," *Stereotact Funct Neurosurg*, vol. 86, no. 2, pp. 113-119, 2008.

[27] Y. Han and H. Park, "Automatic brain MR image registration based on Talairach reference system," in *Proc. ICIP*, 2003, pp. 1097-1100.

[28] L. Verard, P. Allain, J. M. Travere, J. C. Baron, and D. Bloyet, "Fully automatic identification of AC and PC landmarks on brain MRI using scene analysis," *IEEE Trans Med Imaging*, vol. 16, no. 5, pp. 610-616, 1997.

[29] K. N. Bhanu Prakash, Q. Hu, A. Aziz, and W. L. Nowinski, "Rapid and automatic localization of the anterior and posterior commissure point landmarks in MR volumetric neuroimages," *Acad Radiol*, vol. 13, no. 1, pp. 36-54, 2006.

[30] G. Zhang, Y. Fu, S. Wang, and W. Gao, "Automatic localization of AC and PC landmarks in T2-weighted MR volumetric neuroimages," in *Proc. ICIA*, 2010, pp. 1830-1834.

[31] B. A. Ardekani and A. H. Bachman, "Model-based automatic detection of the anterior and posterior commissures on MRI scans," *Neuroimage*, vol. 46, no. 3, pp. 677-682, 2009.

[32] S. Pallavaram, H. Yu, P.-F. D'Haese, J. Spooner, T. Koyama, B. Bodenheimer, C. Kao, P. E. Konrad, and B. M. Dawant, "Automated selection of anterior and posterior commissures based on a deformable atlas and its evaluation based on manual selections by neurosurgeons," in *Proc. SPIE Med. Imaging*, 2007, pp. 65091C-65091C.

[33] P. Anbazhagan, A. Carass, P. L. Bazin, J. L. Prince, "Automatic estimation of midsagittal plane and AC-PC alignment on nonrigid registration," in *Proc. ISBI*, 2006, pp. 828-831.

[34] Y. Liu, R. T. Collins, and W. E. Rothfus, "Robust midsagittal plane extraction from normal and pathological 3-D neuroradiology image," *IEEE Trans Med Imaging*, vol. 20, no. 3, pp. 173-192, 2001.

[35] L. Teverovskiy and Y. Liu, "Truly 3D midsagittal plane extraction for robust neuroimage registration," in *Proc. ISBI*, 2006, pp. 860-863.

[36] M. E. Brummer, "Hough transform detection of the longitudinal fissure in tomographic head images," *IEEE Trans Med Imaging*, vol. 10, no. 1, pp. 74-81, 1991.

[37] I. Volkau, K. B. Prakash, A. Ananthasubramaniam, A. Aziz, and W. L. Nowinski, "Extraction of the midsagittal plane from morphological neuroimages using the kullback-leibler's measure," *Mel Image Anal*, vol. 10, no. 6, pp. 863-874, 2006.

[38] M. M. Chakravarty, *et al.*, "The creation of a brain atlas for image guided neurosurgery using serial histological data," *Neuroimage*, vol. 30, no. 2, pp. 359-376, 2006.

[39] J. Yelnik, *et al.*, "A three-dimensional, histological and deformable atlas of the human basal ganglia. I. Atlas construction based on immunohistochemical and MRI data," *Neuroimage*, vol. 34, no. 2, pp. 618-638, 2007.

[40] P.-F. D'Haese, E. Cetinkaya, P. E. Konrad, C. Kao, and B. M. Dawant, "Computer-aided placement of deep brain stimulators: from planning to intraoperative guidance," *IEEE Trans Med Imaging*, vol. 24, no. 11, pp. 1469-1478, 2005.

[41] W. L. Nowinski, D. Belov, P. Pollak, and A. L. Benabid, "Statistical analysis of 168 bilateral subthalamic nucleus implantations by means of the probabilistic functional atlas," *Neurosurgery*, vol. 57, no. 4, pp. 319-330, 2005.

[42] T. Guo, A. G. Parrent, and T. M. Peters, "Surgical targeting accuracy analysis of six methods for subthalamic nucleus deep brain stimulation," *Comput Aided Surg*, vol. 12, no. 6, pp. 325-334, 2007.

[43] S. Pallavaram, P.-F. D'Haese, W. Lake, P. E. Konrad, B. M. Dawant, and J. S. Neimat, "Fully automated targeting using nonrigid image registration matches accuracy and exceeds precision of best manual approaches to subthalamic deep brain stimulation targeting in Parkinson's disease," *Neurosurgery*, vol. 76, no. 6, pp. 756-765, 2015.

[44] P-F. D'Haese, S. Pallavaram, C. Kao, J. S. Neimat, P. E. Konrad, and B. M. Dawant, "Effect of data normalization on the creation of neuro-probabilistic atlases," *Stereotact Funct Neurosurg*, vol. 91, no. 3, pp. 148-152, 2013.

[45] M. M. Chakravarty, A. F. Sadikot, J. Germann, P. Hellier, G. Bertrand, and D. L. Collins, "Comparison of piece-wise linear, linear, and nonlinear atlas-to-patient warping techniques: analysis of the labeling of subcortical nuclei for functional neurosurgical applications," *Hum Brain Mapp*, vol. 30, no. 11, pp. 3574-3595, 2009.

[46] F. Castro, C. Pollo, R. Meuli, P. Maeder, O. Cuisenaire, M. B. Cuadra, J-G. Villemure, and J-P. Thiran, "A cross validation study of deep brain stimulation targeting: from experts to atlas-based, segmentation-based and automatic registration algorithms," *IEEE Trans Med Imaging*, vol. 25, no. 11, pp. 1440-1450, 2006.

[47] C. Lindner, S. Thiagarajah, J. M. Wilkinson, T. A. Consortium, G. A. Wallis, and T. F. Cootes, "Fully automatic segmentation of the proximal femur using random forest regression voting," *IEEE Trans Med Imaging*, vol. 32, no. 8, pp. 181-189, 2013.

[48] B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu, "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in *Proc. MICCAI*, 2012, pp. 590-598.

[49] R. L. O'Gorman, *et al.*, "Optimal MRI methods for direct stereotactic targeting of the subthalamic nucleus and globus pallidus," *Eur Radiol*, vol. 21, no. 1, pp. 130-136, 2011.

[50] C. R. Traynor, G. J. Barker, W. R. Crum, S. C. Williams, and M. P. Richardson, "Segmentation of the thalamus in MRI based on T1 and T2," *Neuroimage*, vol. 56, no. 3, pp. 939-950, 2011.

[51] M. R. Wiegell, D. S. Tuch, H. BW Larsson, and V. J. Wedeen, "Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging," *NeuroImage*, vol. 19, no. 2, pp. 391–401, 2003.

[52] J. Glaister, A. Carass, J. V. Stough, P. A. Calabresi, and J. L. Prince, "Thalamus parcellation using multi-modal feature classification and thalamic nuclei priors," in *Proc. SPIE Medical Imaging*, 2016, pp. 97843J-97843J.

[53] T. EJ Behrens, *et al.*, "Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging," *Nat Neurosci*, vol. 6, no. 7, pp. 750–757, 2003.

[54] H. Johansen-Berg, *et al.*, "Functional–anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus," *Cereb Cortex*, vol. 15, no. 1, pp. 31-39, 2005.

[55] D. Zhang, A. Z. Snyder, J. S. Shimony, M. D. Fox, and M. E. Raichle, "Noninvasive functional and structural connectivity mapping of the human thalamocortical system," *Cereb Cortex*, vol. 20, no. 5, pp. 1187–1194, 2010.

[56] A. T. Newton, B. M. Dawant, and P. D'Haese, "Visualizing intrathalamic structures with combined use of MPRAGE and SWI at 7T," in *Proc. ISMRM*, 2014, pp. 122.

[57] B. T. Wang, S. Poirier, T. Guo, A. G. Parrent, T. M. Peters, and A. R. Khan, "Generation and evaluation of an ultra-high-field atlas with applications in DBS planning," in *Proc. SPIE Medical Imaging*, 2016, pp. 97840H-97840H.

[58] J. Kim, Y. Duchin, G. Sapiro, J. Vitek, and N. Harel, "Clinical deep brain stimulation region prediction using regression forests from high-field MRI," in *Proc. ICIP*, 2015, pp. 2480-2484.

CHAPTER II

MULTI-SURGEON, MULTI-SITE VALIDATION OF A TRAJECTORY PLANNING ALGORITHM FOR DEEP BRAIN STIMULATION PROCEDURES[1]

Yuan Liu[1], Peter E. Konrad[2], Joseph S. Neimat[2], Stephen B. Tatter[3], Hong Yu[4], Ryan D. Datteri[1], Bennett A. Landman[1], Jack H. Noble[1], Srivatsan Pallavaram[1], Benoit M. Dawant[1], and Pierre-François D'Haese[1]

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37232 USA

[2]Department of Neurosurgery, Vanderbilt University, Nashville, TN 37235 USA

[3]Department of Neurosurgery, Wake Forest University, Winston-Salem, NC 27157 USA

[4]Department of Neurosurgery, Stanford University, Stanford, CA 94305 USA

[1] [This chapter has been published in *IEEE Transactions on Biomedical Engineering*, Volume 61, Issue 9, pp. 2479-2487, September 2014.]

Abstract

Deep brain stimulation, which is used to treat various neurological disorders, involves implanting a permanent electrode into precise targets deep in the brain. Reaching these targets safely is difficult because surgeons have to plan trajectories that avoid critical structures and reach targets within specific angles. A number of systems have been proposed to assist surgeons in this task. These typically involve formulating constraints as cost terms, weighting them by surgical importance, and searching for optimal trajectories, in which constraints and their weights reflect local practice. Assessing the performance of such systems is challenging because of the lack of a ground truth and clear consensus on an optimal approach among surgeons. Due to difficulties in coordinating inter-institution evaluation studies, these have been performed so far at the sites at which the systems are developed. Whether or not a scheme developed at one site can also be used at another is thus unknown. In this article, we conduct a study that involves four surgeons at three institutions to determine whether or not constraints and their associated weights can be used across institutions. Through a series of experiments, we show that a single set of weights performs well for all surgeons in our group. Out of 60 trajectories, our trajectories were accepted by a majority of neurosurgeons in 95% of the cases and the average acceptance rate was 90%. This study suggests, albeit on a limited number of surgeons, that the same system can be used to provide assistance across multiple sites and surgeons.

2.1. Introduction

Deep brain stimulation (DBS) procedures, which are widely used to treat patients suffering from neurological disorders such as Parkinson's disease, tremor, or dystonia [1], require placing a permanent four contact electrode in specific deep brain nuclei. Multiple studies have indicated

that precise planning of the trajectory is necessary to avoid psychological and motor side effects or complications such as hemorrhage. Comprehensive reviews of risks associated with the procedure can be found in [2-3]; the latter reports a risk of hemorrhage ranging from 0.3% to 10% per patient, depending on the site at which the procedure is performed. At leading clinical sites, a 1% risk of intracranial hemorrhage associated with permanent neurological deficit is currently considered to be the norm. Planning the procedure is thus a task that requires expertise, and a review of the literature shows that a number of approaches have been proposed over the years to assist surgeons in selecting trajectories for neurosurgery in general and for DBS procedures in particular. In 1997, Vaillant *et al.* [4] quantified the risk of a candidate trajectory using a weighted sum of the voxel costs manually assigned according to the tissue significance along a trajectory. A decade later, Brunenberg *et al*. [5] approached the problem again. They evaluated the risk of an automatic trajectory using the maximum voxel cost, defined as the minimum distance to vessels and ventricles. The set of all possible trajectories that satisfied the cost criteria defined as distance thresholds to each structure was then displayed to surgeons for them to choose from. This was followed by the work of Navkar *et al*. [6], who attempted to facilitate the path selection process by overlaying color-coded risk maps on a rendered surface of the patient's head. Shamir *et al*. [7] proposed to account for both the weighted sum of voxel costs and the maximum voxel cost with respect to vessels and ventricles along the trajectory, and combined the individual costs of each structure with risk levels defined by surgeons. In a more comprehensive framework, Essert *et al*. [8] formalized various surgical rules as separate geometric constraints and aggregated them into a global cost function for path optimization. Following a similar idea, Bériault *et al*. [9] defined constraints using multi-modality images and modeled the trajectory as a cylinder.

While these methods suggest the feasibility of developing reliable computer-assisted planning systems, assessing their performance remains challenging, as no universal ground truth exists. Qualitative evaluations have involved user-experience questionnaire about effectiveness of the interactive path selection software by neurosurgeons [5-6]. Recently, Shamir *et al*. [7] and Essert *et al*. [8] have quantitatively compared the safety of automatic trajectories with manual ones using their distances to critical structures. Essert *et al*. [8] also reported the global and individual costs between those two types of trajectories on 30 cases, which, as stated in the paper, may favor the automatic approach that is designed to minimize these cost values. Bériault *et al*. [9] performed not only a quantitative evaluation similar to others [7-8], but also a qualitative validation where the automatic trajectories were rated against one set of manual trajectories by two neurosurgeons for 14 cases retrospectively. More recently, Bériault *et al*. [10] tested their method prospectively on 8 cases in a study where the system proposed five trajectories in the first round. If none of these were found acceptable, the system was initialized interactively by surgeons up to three times to compute a local optimal trajectory each time. The surgeon could then choose either one of the system-generated trajectories or a trajectory selected manually in the normal delivery of care. Out of these 8 cases on which this method was evaluated, one of the five initial system-generated solutions was selected for five cases, a manually initialized but automatically computed solution was selected for one case, one case was planned manually by the surgeon and for the last case both solutions were deemed equivalent.

Importantly, even though all the current automatic trajectory computation algorithms involve a cost function with multiple terms modeling surgical constraints and the selection of weights for each of these terms, no study has explored whether or not individual surgeon preferences would necessitate the adjustment of weights or even the constraints. It is thus not

known whether one could develop a system that could be used by a multitude of surgeons at multiple sites or if some type of training would be required when a system is fielded at a new site. In this article, we report on a number of experiments we have performed as an extension of previous work [11] to begin answering these questions both retrospectively and pseudo-prospectively. In the rest of the article, we first provide details about the data, the cost function, and the training method we have used. The various experiments we have performed to study whether or not one could capture surgeon preferences and if these were of clinical significance are described next. This is followed by a description of pseudo-prospective studies we have performed with 3 experienced movement disorder surgeons (one affiliated with our institution, one with Wake Forrest University, and one with Stanford University) who collectively perform in excess of 200 cases a year. We subsequently present the results we have obtained for each of these experiments. These results are discussed and avenues for future investigations are presented in the last section of this article.

2.2. Methods

2.2.1. Data

Thirty DBS patients with bilateral implantations in the subthalamic nucleus (STN) are included in this study, for a total of 60 distinct trajectories. All subjects provided informed consent prior to enrollment. For each subject, our dataset includes magnetic resonance (MR) T1-weighted (T1-w) images without and with contrast agent (MR T1(-C)), and the target position at which the implant has been surgically placed. The MRIs (TR: 7.9 ms, TE: 3.65 ms, $256 \times 256 \times 170$ voxels, with voxel resolution $1 \times 1 \times 1$ mm^3) were acquired using the SENSE parallel imaging technique

from Philips on a 3T scanner. The MR T1-C is rigidly registered to the MR T1-w prior to trajectory optimization with a standard intensity-based technique that uses Mutual Information [12-13] as its similarity measure.

2.2.2. Trajectory Planning Approach

Our approach involves formulating surgical constraints as individual cost terms and finding the optimal trajectory in a limited search space based on a cost function that aggregates those terms.

2.2.2.1. Formulation of Surgical Constraints

We gathered a number of constraints expressed as rules through extensive discussions with the two experienced movement disorder neurosurgeons affiliated with our institution. These rules are based on factors including safety, esthetic, and lead orientation with respect to the target structures and are similar but not identical to constraints used by others. For example, constraints such as minimizing the path length [8], minimizing overlap with caudate [9] and maximizing the orientation of the electrode depending on target shape [8], were not considered significant by our surgeons and thus not included in our system. Some of these rules cannot be violated while others lead to trajectories that are less desirable when violated. Using these rules we designed a cost function that contains hard constraints, i.e., constraints that cannot be violated, and soft constraints, i.e., constraints that penalize the trajectories more as the extent to which they are violated increases.

1) Rule 1 (Burr hole): The entry point should be posterior to the hairline for cosmetic reasons and anterior to the motor cortex to avoid side effects.

This is incorporated in our algorithm by limiting the generation of candidate trajectories within an allowable bounding box defined on a skin-air interface mesh.

2) Rule 2 (Vessels): The trajectory should stay away from vessels to minimize the risk of hemorrhage.

This is modeled as a soft constraint that penalizes trajectories that are closer than 3 mm from a vessel, with the corresponding cost term f_2 defined as follows:

$$f_2 = \begin{cases} (3 - d_2)/(3 - d_{2_{\min}}) & \text{if } d_2 < 3 \\ 0 & \text{if } d_2 \geq 3 \end{cases} \quad (2.1)$$

where d_2 is the Euclidean distance from the trajectory to the vessels and $d_{2_{\min}}$ is the minimum value of d_2 over the set of trajectories being evaluated for one case. A hard constraint is also defined which sets the total cost of a trajectory to be maximum if it penetrates a vessel.

3) Rule 3 (Ventricle): Ventricles are structures that should not be perforated by the trajectory, while it is desirable to stay close to them to favor approaches that maximize the coverage of the region of interest in the STN.

This is modeled through both a soft constraint that penalizes trajectories that are further than 2 mm to the ventricles, and a hard constraint that sets the total cost of a trajectory to be maximum if it is closer than 2 mm. The corresponding cost term f_3 for the soft constraint is defined as follows:

$$f_3 = (d_3 - 2)/(d_{3_{\max}} - 2)\text{, if } d_3 \geq 2, \quad (2.2)$$

where d_3 is the Euclidean distance to the ventricles and $d_{3_{\max}}$ is the maximum value of d_3 over the set of trajectories being evaluated for one case.

4) Rule 4 (Sulci): It is unsafe to cross sulci because small vessels that are invisible on preoperative imaging are often present at the base of the sulci.

This is modeled as a soft constraint that penalizes trajectories that intersect the cortical surface at the base of the sulci. The cost term for this constraint, f_4, is defined as follows:

$$f_4 = (d_{4_{\max}} - d_4)/(d_{4_{\max}} - d_{4_{\min}}) \tag{2.3}$$

where d_4 is a distance quantity to the sulci, and $d_{4_{\min}}$, $d_{4_{\max}}$ are the minimum and maximum value of d_4 over the set of trajectories being evaluated respectively. The search area defined by the bounding box in *Rule* 1 is large enough to cover both sulcal and gyral regions, so that trajectories with f_4 approaching 1 intersect with sulci and trajectories with f_4 approaching 0 intersect with gyri. Further details on the computation of d_4 are provided later in this section.

5) Rule 5 (Suture): Some neurosurgeons prefer the entry point to be near the coronal suture.

This is modeled as a soft constraint that penalizes entry points that are located posteriorly or anteriorly from the middle of our allowable entry region, which is used as an approximation of the coronal suture. The cost term for this constraint, f_5, is defined as follows:

$$f_5 = d_5/d_{5_{\max}} \tag{2.4}$$

where d_5 is the Euclidean distance from the entry point of a trajectory to the suture, and $d_{5_{\max}}$ is the maximum value of d_5 among all trajectories being evaluated.

6) Rule 6 (Thalamus): Intersecting the lateral edge of the thalamus is desired by some surgeons when targeting the STN because it can be used as an electrophysiological landmark before reaching the target (this was expressed as a desirable trajectory property by one of the two surgeons at our institution).

This is modeled as a soft constraint that penalizes trajectories passing through the thalamus for less than one millimeter, with the corresponding cost f_6 defined as follows:

$$f_6 = \begin{cases} 1 - d_6 & \text{if } d_6 < 1 \\ 0 & \text{if } d_6 \geq 1 \end{cases}, \tag{2.5}$$

where d_6 is the length of the trajectory passing through the thalamus.

Cost function: The overall cost function, f_{total}, is a weighted sum of the individual cost terms

associated with each of the soft constraints 2-6 we have defined. It also contains the hard constraints that are part of *Rule 2* and *Rule 3*:

$$f_{\text{total}} = \begin{cases} \sum_{i=2}^{6} w_i f_i & \text{if } d_2 \geq 0 \cap d_3 \geq 2 \\ 1 & \text{otherwise} \end{cases}, \text{s.t. } \{w_i\} \geq 0, \sum_{i=2}^{6} w_i = 1 \qquad (2.6)$$

where $\{d_i\}$ are the distance quantities discussed above, and $\{w_i\}$ are the values used to weigh the relative importance of each cost term. The method we use to obtain those weights is described in detail in the next section. All cost terms $\{f_i\}$ are normalized, and the weights $\{w_i\}$ are forced to be positive and their sum equal to 1. The overall cost f_{total} thus ranges from 0 to 1. If a trajectory hits a vessel, i.e., when $d_2 < 0$, or if a trajectory is closer than 2 mm to the ventricles, i.e., when $d_3 < 2$, then f_{total} is set to its maximum value. Otherwise, f_{total} is equal to the weighted sum of the cost terms.

2.2.2.2. Path Optimization

The path optimization process is an exhaustive search that consists of generating all candidate trajectories in a specific search space, computing the cost for each trajectory based on the cost function described above and finding the trajectory with the lowest cost, as illustrated in Figure 2.1.

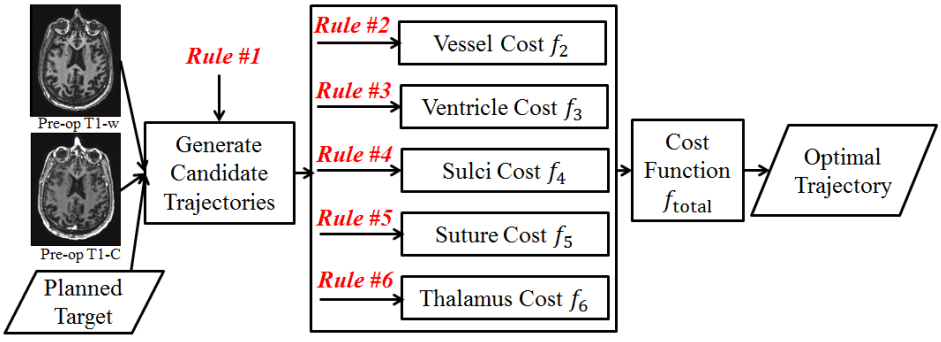

Figure 2.1. Overview of the path optimization process.

Path Generation

Each trajectory is modeled as a linear segment with a common target point and can be uniquely represented by its entry point located on the skin-air interface. This interface is a dense triangulated mesh with vertices representing possible entry points, and is obtained automatically by extracting the isosurface of the patient's T1-w image after smoothing using marching cubes algorithm [14] and then filtering out extraneous components.

The allowable entry region specified by *Rule 1* is determined by projecting 4 points that have been selected manually by one neurosurgeon on one T1-w reference volume onto the patient's head surface. This reference volume is registered to the corresponding patient volume using non-linear registration and the transformation that registers these volumes is used to project the points. Those four points defines a tetrahedron that bounds the region, and all vertices within this region on the skin-air interface are considered to be candidates. This set of candidate entry points typically consists of 2000 points separated by an average distance of 1 mm and serves as the constrained search space in which the optimal solution is found.

Cost Computation

Several rules (2, 3, 4 and 6) involve anatomical structures therefore their segmentations are required to compute the cost function. Vessels were identified by thresholding the difference image created by subtracting the MR T1-w from co-aligned T1-C within the skull region. Ventricles and thalamus were segmented by combining atlas-based segmentations from four atlases using STAPLE [15]. From those segmentations distance maps were computed with a fast marching method algorithm [16]. These were then used to calculate the distance between a trajectory and the various structures using the technique described by Noble *et al*. [17]. All nonrigid registrations were performed with the Adaptive Bases Algorithm [18].

The distance to the sulci, d_4, is computed in several steps, as illustrated in Figure 2.2. First, the cortical surface is extracted using the LongCruise tool [19] (represented schematically by the colored curve). Then the deepest intersection point of the trajectory with this surface is found (shown as a star in the figure). Vertices on the surface mesh that fall within N neighbor edge connections from this point are identified (indicated by dots in the same figure) and projected onto the trajectory to compute a signed distance from the intersection point to the projected points (positive inside the cortex, negative outside). The distance d_4 is a weighted average of those distance values, where the weights are defined to be the reciprocal of the number of points with the same neighbor edge connections so that points with approximately equivalent geodesic distance to the intersection point contribute equally. As can be seen in Figure 2.2, d_4 is negative when intersecting a valley (i.e., a sulcus), is around zero when intersecting neither a peak nor valley, and is positive when intersecting a peak (i.e., a gyrus). Hence f_4 is minimized when a trajectory intersects the cortical surface through the top of a gyrus.

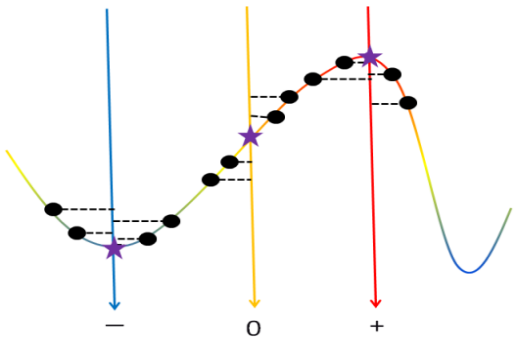

Figure 2.2. A 2D illustration of d_4. The curve represents the cortical surface and the lines represent candidate trajectories, all color coded by this sulci distance, d_4, where blue and red indicate lower and higher values respectively. The intersection points are indicated by stars and neighbor points by dots. From left to right, the average projection distance for each trajectory is negative, nearly zero, and positive, respectively.

2.3. Experiments

We conducted a set of experiments, first retrospectively and internally with two surgeons at our institution and then pseudo-prospectively with two additional surgeons at external institutions. In this article, we define a pseudo-prospective experiment to be the selection of a trajectory as it would be done in clinical practice but without actually performing the procedure (more details on the way the experiments were performed are provided below). All experiments have been performed with the research version of the FDA cleared surgical planning system used clinically by all four neurosurgeons called WayPoint Navigator [20] that is distributed by FHC, Inc. Bowdoin, ME, USA.

2.3.1. Retrospective Internal Studies

The retrospective studies were performed by the two co-author neurosurgeons affiliated with our institution and subsequently referred to as surgeons A and B. Surgeon-specific weighting parameters were learned prior to any evaluation experiments using two sets of training volumes, one set per surgeon. To do this, we collected 10 pair of images (T1-w and T1-C) of patients operated on by surgeon A and another 10 operated on by surgeon B, all with bilateral implantation. For each of these volumes, the planned trajectories of both left and right sides selected by the surgeons in the normal delivery of care were available. These images and trajectories constitute the training sets, leading to 20 training trajectories from 10 patients per surgeon. Surgeon-dependent weights for our algorithm were determined heuristically and iteratively as follows. First, initial weight values were set according to the relative importance of the constraint, as stated by the surgeon. Then, automatically generated and manually selected trajectories were presented simultaneously in the surgical planning system to the surgeons who

were blind to the way the trajectories were generated. The surgeons were asked to compare the trajectories using a four level comparative scale. If the trajectory being evaluated was chosen, i.e., if it better matched the surgeon's preferences than the one chosen manually, it was classified as "Excellent". If the surgeon was not able to determine which one was superior, e.g., if they were too close to each other, it was rated as "Equivalent". If the trajectory under evaluation was not chosen but could still be used clinically, i.e., it was considered to be safe, it was classified as "Acceptable", and otherwise was classified as "Rejected". A computed trajectory was considered to be successful if rated "Acceptable" or better. Based on this surgeon's feedback, weights were manually adjusted and the experiment repeated until each surgeon ranked all his training trajectories as at least acceptable, i.e., safe. This required four iterations. Once these parameters were estimated, they were fixed and used for all evaluation experiments.

To test our method with weights obtained in the training phase, we acquired another 10 image sets of patients operated on by surgeon B, all with bilateral implantation, for which the planned trajectories selected by surgeon B in the normal delivery of care were available. To obtain manual trajectories on the same set of image volumes, surgeon A was asked to select a trajectory manually on both sides of each of these volumes as if performing the procedure. These 10 volumes as well as both surgeons' manual trajectories constitute the testing set, leading to 20 testing trajectories per surgeon. These manual trajectories were used in the various experiments described below. Evaluation of a trajectory in these experiments is achieved as described in the training session, i.e., the surgeons were asked to rate the trajectory against their own manual one without knowing the trajectories' provenance with the four-level comparative rating scale described earlier.

The goals of our retrospective validation experiments were threefold. First, test whether

or not we could estimate surgeon-specific parameter values that would lead to acceptable trajectories for the surgeon for whom the parameters were tuned. Second, evaluate whether or not surgeon-specific weighting parameters were indeed necessary. Third, test whether or not one surgeon would prefer trajectories generated automatically or by another surgeon, i.e., would automatic trajectories appear any different to a surgeon than trajectories selected manually by another experienced surgeon. The following experiments were performed to answer these questions.

1) Experiment 1: evaluation of the trajectories computed based on surgeons' own preferences

The goal of the first experiment is to assess the degree to which the weighting factors optimized for each neurosurgeon effectively capture their preferences. The surgeon is asked to evaluate the trajectories computed by our method using his own preferences against his own clinical ones.

2) Experiment 2: evaluation of the trajectories computed based on other surgeons' preferences

The goal of the second experiment is to assess the benefit of using surgeon-specific parameters. The surgeon is asked to evaluate the trajectories computed by our method using the weight values estimated for the other surgeon against his own clinical ones.

Results of experiment 1 and 2, detailed in the next section, led us to the conclusion that the set of weights optimized for surgeon B, could be used to pursue our validation study. Thus the computed trajectories evaluated in the following experiments are all generated by this set of weights.

3) Experiment 3: evaluation of the trajectories manually picked by other surgeons

The goal of the third experiment is to assess if automatically generated trajectories are distinguishable from trajectories manually selected by another surgeon. Each surgeon was

presented with a choice between his trajectory and the trajectory manually selected by the other surgeon. The results of this experiment are compared to the results of the experiments in which each surgeon is presented with a choice between the automatically generated trajectory and his manual trajectory. The automatically generated trajectories for surgeon A are the same as those generated for him in experiment 2 and trajectories computed for surgeon B are the same as those evaluated in experiment 1, because, as discussed above, the weight values optimized for surgeon B were used for both surgeons in this experiment.

2.3.2. Pseudo-prospective Studies

We define the pseudo-prospective scenario as putting the surgeons in the same conditions as a prospective scenario but using retrospective cases, i.e., we ask the surgeon to redo plans on patients that already underwent the surgery. Because in this study we focus on evaluating the trajectories, we fix the target point to be the clinical target point chosen during the real plan. To be more specific, for each case, surgeons were provided with a single trajectory. If they thought the trajectory was adequate to be used clinically, it was labeled as "Acceptable"; otherwise, it was labeled as "Rejected". In this case, the surgeons would adjust it to make it useable and provide cause(s) for rejection. The two surgeons at our site evaluated and interacted with the trajectories on a laptop running the planning system as they would clinically. This typically involves checking the trajectory along its length in orthogonal views and in a "bird's eye view", i.e., in slices that are reformatted to be perpendicular to the trajectory. Surgeons were allowed to modify the trajectory to check whether or not a better one could be found. The surgeons at remote sites followed the same process but the computer was accessed remotely and its desktop shared. Although the goal of our pseudo-prospective study is to test in clinical conditions

whether or not our algorithm could be used to assist the surgeon in selecting an optimal trajectory, we first investigate whether or not surgeons behave differently when choosing between two trajectories or when deciding to accept or reject a single trajectory. We do this by presenting surgeons A and B with the automatic trajectories they were presented in experiment 2 for surgeon A and in experiment 1 for surgeon B. We then extend our analysis to the multi-site evaluation using a larger dataset.

4) Experiment 4: evaluation of the computed trajectories in pseudo-clinical setting within internal site

In this experiment, we put surgeons A and B into clinical condition to test the automatic trajectories computed with surgeon B's weights. We present each surgeon with 20 trajectories computed for the volumes in the testing set and ask them to either accept or reject the trajectories. These trajectories are the same as those presented to surgeon A in experiment 2 and surgeon B in experiment 1.

5) Experiment 5: evaluation of the computed trajectories in pseudo-clinical setting among multi-institutional surgeons

In this last experiment, we tested the automatic algorithm on the complete set of patients, i.e., on 60 trajectories. These trajectories were evaluated by one surgeon affiliated with Vanderbilt University, one with Stanford University, and one with Wake Forest University. Because the weights used to generate these trajectories were estimated to capture surgeon B's preferences, he was eliminated from this pseudo-prospective experiment to avoid any bias.

As several constraints are based on the distance to specific structures, their under- or over-segmentation could affect our results. To avoid this potential confounding factor and because the focus of this paper is not on the validation of the segmentation algorithms, prior to

experiment 5, we verified the segmentations of structures including ventricles and veins around the ventricles and refined them if needed.

2.4. Results

Parameter selection was carried out internally for surgeons A and B. For surgeon A, weighting parameters were chosen through the training process to be 0.18 for f_2, 0.18 for f_3, 0.35 for f_4, 0.24 for f_5, 0.05 for f_6. For surgeon B, weighting parameters were 0.20 for f_2, 0.27 for f_3, 0.40 for f_4, 0.13 for f_5, 0.00 for f_6. For illustration purposes, Figure 2.3 shows for one case color maps that convey the costs associated with entry points within the allowable search region (blue = low cost, red = high cost). Panels (a)-(e) show individual costs and (f) shows the overall cost. Figure 2.4 shows a manual trajectory (green) and two computed trajectories rated against this manual one for one case, with one rated as excellent (blue), and the other rejected (red). The rejected trajectory is closer to the sulcus, which is the reason for rejection. The trajectory rated as excellent is further away from vessels compared to the manual trajectory.

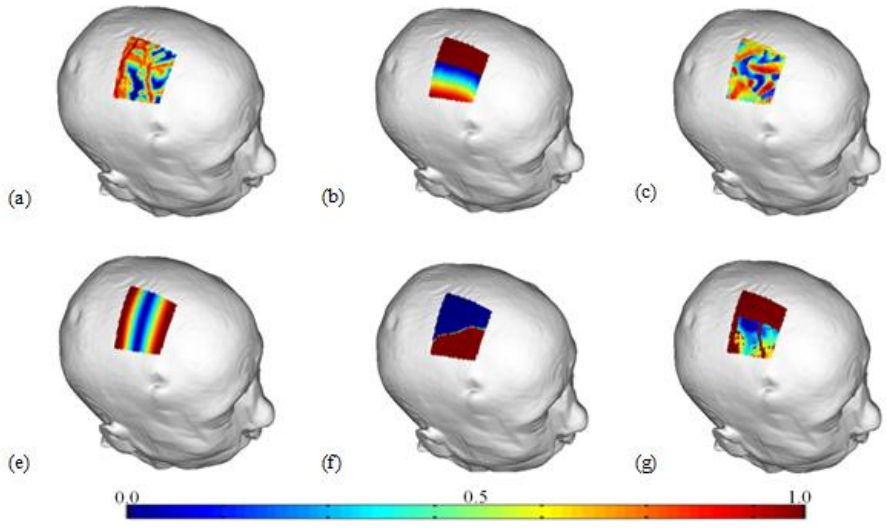

Figure 2.3. Color maps showing the cost for (a) f_2, the vessel cost; (b) f_3,the ventricle cost; (c) f_4, the sulci cost; (d) f_5, the suture cost; (e) f_6, the thalamus cost; (f) f_{total}, the total cost. (blue = low cost, red = high cost)

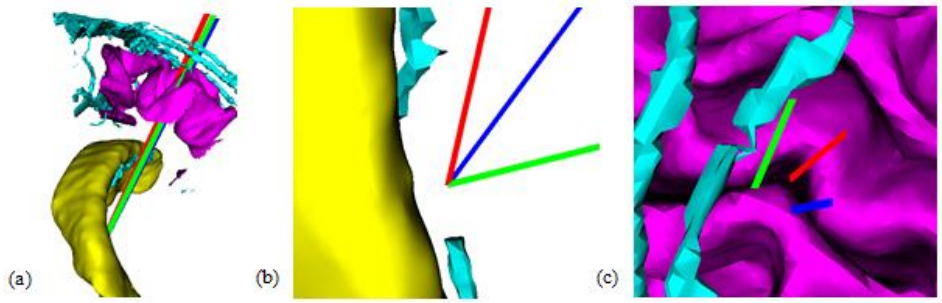

Figure 2.4. Manual trajectory (green), trajectory rated as excellent (blue), and trajectory rated as rejected (red), both being rated against the manual one. The surrounding structures include the ventricles (yellow), vessels (cyan), and the cortical surface (magenta). (a) a global view; (b) zoomed in view around the ventricular surface; (c) zoomed-in view around the cortical surface.

Results for experiment 1 in which we assess the degree to which these weighting factors effectively capture the neurosurgeon's preferences are presented in Figure 2.5. For surgeon A, 25% of the 20 available test cases are rated as excellent, 50% as equivalent, 15% as acceptable, and 10% as rejected, while for surgeon B, 25% are rated as excellent, 10% as equivalent, 50% as acceptable, and 15% as rejected. Acceptance rates of 90% and 85% for surgeons A and B suggest that the surgeon's preferences were reasonably captured in the weighting parameters.

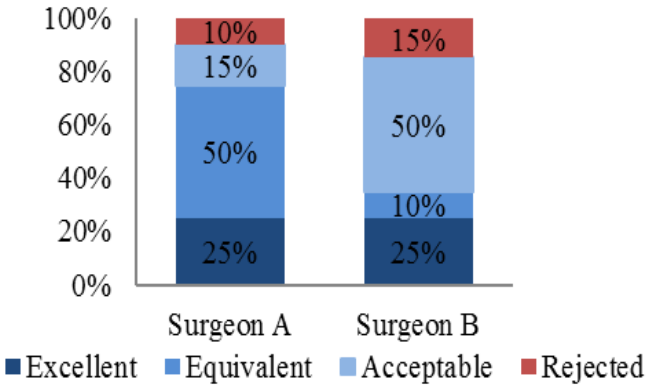

Figure 2.5. Rating distribution of the trajectories computed using the surgeon's own set of weights for surgeons A (left) and B (right). Red represents the percentage of cases being rejected and light, medium, dark blue represents the percentage of cases rated as acceptable, equivalent, and excellent respectively.

Results for experiment 2 in which we assess the benefit of using surgeon's specific

parameters are shown in Figure 2.6. This figure shows the results of the rating experiments conducted with surgeons A (left) and B (right). The light blue columns represent rating scores for the trajectories computed using the same surgeon's set of weights and dark red for those computed using the other surgeon's set of weights. For each surgeon, 15 out of the overall 20 cases ended up having the same rating scores between the trajectories computed using the two sets of weights. Only cases when these two rating scores were different are shown in this figure. The left panel shows that surgeon A consistently prefers trajectories computed using surgeon B's weights and rated all of those trajectories better than the ones computed using his own weights. In the right panel, the plot shows that surgeon B rejected several trajectories computed using surgeon A's weights and in general prefers his own set of weights. This suggests that the set of weighting parameters chosen for surgeon B is acceptable for both surgeons. Results obtained with this experiment led us to continue our analysis with only the set of weights optimized for surgeon B.

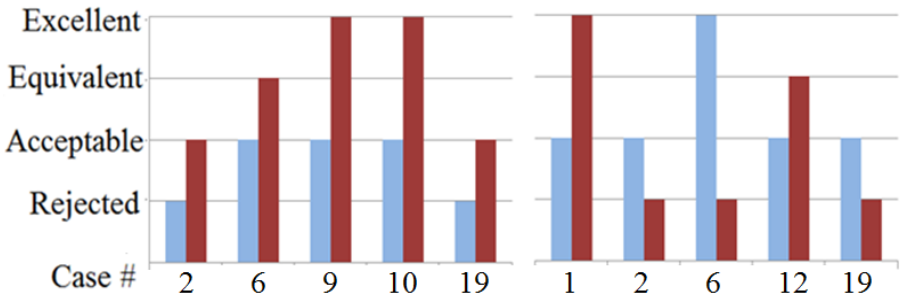

Figure 2.6. Pairwise comparison of rating scores. Left: experiment conducted by surgeon A; right: experiment conducted by surgeon B. Light blue represents the rating for the trajectory computed using the same surgeon's set of weights and dark red indicates the rating for the trajectory computed using the other surgeon's set of weights. Only cases that resulted in different rating between the two trajectories are shown.

The results of experiment 3 in which we assess whether or not a surgeon has a preference for automatically generated trajectories over trajectories selected manually by another surgeon

are shown in Figure 2.7. Results obtained with surgeon A are plotted in the left panel and those obtained with surgeon B in the right panel. For each surgeon we show the distribution of ratings for the other surgeon's manual trajectories on the left and the distribution of ratings for the trajectories computed using surgeon B's weights on the right. This figure shows that surgeon A tends to prefer the automatic trajectories over surgeon B's manual trajectories and surgeon B rejected fewer automatic trajectories than trajectories produced manually by surgeon A. As is the case for our pseudo-prospective study (see below), the most frequent reason for rejecting an automatic trajetory is its position relative to the sulci.

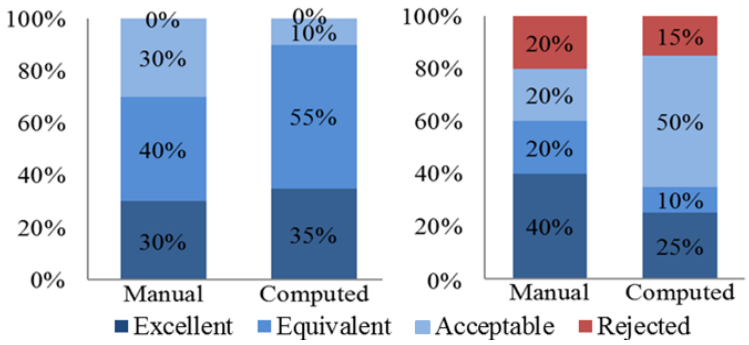

Figure 2.7. Rating distributions of trajectories evaluated by surgeons A (left) and B (right). For each surgeon the ratings of the other surgeon's manual trajectories are shown in the left and those of computed trajectories are shown in the right. Red represents the percentage of cases being rejected and light, medium, dark blue represents the percentage of cases rated as acceptable, equivalent, and excellent respectively.

To quantitatively evaluate automatic trajectories against manual ones, we compare vessel, ventricle, sulci, suture, and overall cost values obtained with each approach as well as the distances to the vessels and ventricles as was done in [7-9]. Table 2.1 reports means, standard deviations, minima and maxima for these quantities. To test whether or not differences are statistically significant, we also perform a Wilcoxon paired two-sided signed rank test and report p-values. Results show that automatically generated trajectories are further away from the

vessels, as desired. For the ventricles, the automatic method leads to lower cost values even though the mean distance is lower for surgeon A than for the automatic approach. This is because several of surgeon A's manal trajectories are closer to the ventricles than the 2 mm specified in *Rule* 3. This violates the hard constraint, thus raising the cost. This hard constraint is also violated for one automatic trajectory because in these experiments trajectories were computed with the automatic segmentations but the statistics are computed with the manually validated and edited segmentations. Automatically generated trajectories lead to lower cost for the sulci and lower suture costs for surgeon B. The overall cost is also statistically significantly lower for the automatic approach. However, as noted by Bériault *et al*. [9], these numbers have to be interpreted with care. Indeed, they only show that an automatic method is generally better at finding a minimum in an analytically defined function than a human operator. They do not show whether or not this function truly captures the decision-making process of a trained neurosurgeon. Ultimately, the safety and clinical acceptability of automatically generated trajectories needs to be assessed by asking experienced surgeons to review and rate each of them, which is done in experiments 4 and 5.

Table 2.1. Distances to vessels and ventricles and all the cost terms for trajectories computed by surgeon B's weights and manual trajectories by surgeons A and B

		Vessel distance	Vessel cost	Ventricle distance	Ventricle cost	Sulci cost	Suture cost	Total cost
Automatic	Mean±Std.	4.50±3.86	0.08±0.13	3.80±1.80	0.28±0.29	0.10±0.08	0.22±0.27	0.17±0.29
	[Min, Max]	[0.24, 15.14]	[0.00, 0.68]	[0.46, 7.05]	[0.01, 1.00]	[0.00, 0.26]	[0.01, 1.00]	[0.00, 1.00]
A's Manual	Mean±Std.	3.38±4.13	0.29±0.33	3.69±2.90	0.46±0.36	0.35±0.21	0.16±0.12	0.48±0.40
	[Min, Max]	[-0.19, 15.22]	[0.00, 1.00]	[-3.98, 8.91]	[0.01, 1.00]	[0.11, 0.73]	[0.03, 0.50]	[0.06, 1.00]
B's Manual	Mean±Std.	3.67±4.10	0.23±0.29	4.30±1.62	0.34±0.26	0.24±0.15	0.24±0.15	0.29±0.31
	[Min, Max]	[-0.44, 15.31]	[0.00, 1.00]	[0.46, 7.31]	[0.07, 1.00]	[0.00, 0.56]	[0.00, 0.56]	[0.01, 1.00]
Automatic vs. A's *p*-value		**0.0089**	**0.0081**	0.5372	0.1092	**0.0001**	0.4186	**0.0029**
Automatic vs. B's *p*-value		0.0925	0.0533	0.2111	0.3659	**0.0029**	0.2276	**0.0063**

Top three rows: mean, standard deviation (Std.), minima, and maxima of all the testing cases; bottom two rows: *p*-values of the statistical tests performed between the automatic trajectories and manual trajectories. *P*-values that are smaller than 0.05 are marked in red bold.

In experiment 4, where the surgeons were asked to either accept or reject the trajectory they were presented with, surgeon A rejected 3 trajectories and surgeon B 7. Figure 2.8 shows the rating scores obtained for these rejected trajectories in experiments 1 and 2 when surgeons were evaluating the trajectories against their own manual selections. Over the trajectories that were rejected in this experiment, only 2 were rejected in the retrospective studies.

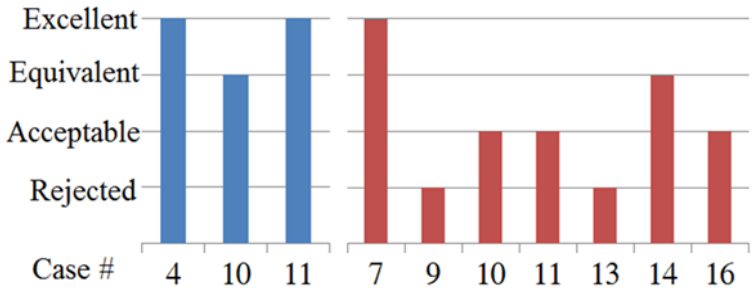

Figure 2.8. Rating scores of the cases rejected by surgeons under the pseudo-prospective scenario, evaluated by surgeons A (left) and B (right).

Table 2.2. Acceptance rate for individual neurosurgeons (A, C, D) and overall in experiment 5

Surgeon A	Surgeon C	Surgeon D	Average
95.00% (57 / 60)	83.33% (50 / 60)	91.67% (55 / 60)	90.00% (54 / 60)

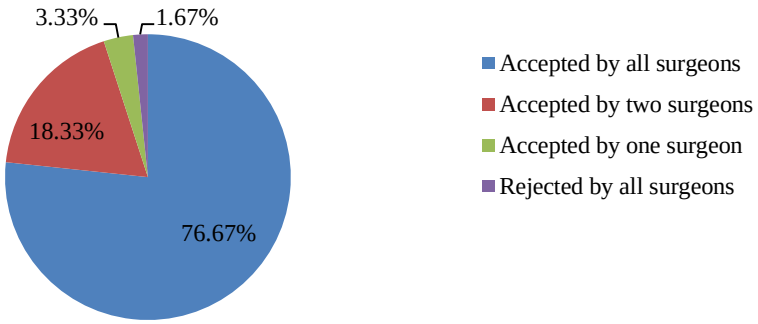

Figure 2.9. Voting distribution according to the trajectory's acceptation/rejection status, i.e., accepted by all, two, one, or rejected by all neurosurgeons, in experiment 5.

Table 2.2 presents the results of experiments 5. It shows that, on average, 90% of the

automatic trajectories were accepted in our pseudo-prospective experiment using 60 trajectories. Figure 2.9 reports the percentage of trajectories that were accepted by all three surgeons, by two surgeons, and by one surgeon as well as the fraction of trajectories rejected by all surgeons.

Figure 2.10 shows the reasons for which the trajectories were rejected and the number of times each surgeon cited a particular reason. The cause for rejection is typically proximity to critical structures, including the sulci, veins/vessels around the ventricles (referred to as ventricular vein), veins/vessels around sulci (referred to as sulcal vein), and veins/vessels around the skull (it may either be in dural matter or in cortical zones, referred to as surface vein), or some other issues (too medial or too lateral). Trajectories rejected for multiple reasons were counted multiple times. As can be seen, the main reason for rejection is related to sulci.

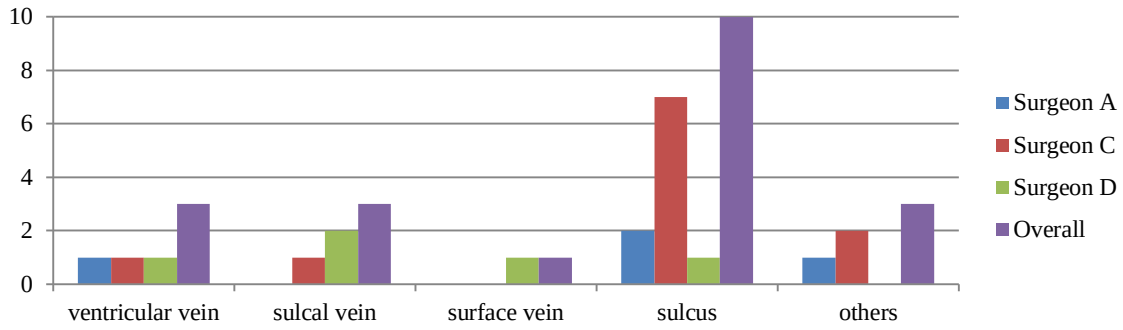

Figure 2.10. Distributions of automatic trajectories rejected by surgeons according to different rejection reasons in experiment 5.

For the trajectories that are rejected, we calculate the angle between the computed trajectories and the ones modified by the surgeons. The average value for this angle is 7.38˚ and its standard deviation is 6.23˚.

2.5. Discussions and Conclusions

The selection of safe trajectories is an important component of DBS procedures and one that requires considerable expertise. In this article, we build on our own and the work of others to investigate whether or not computer-assistance would be valuable and whether or not systems need to be adapted for each surgeon to capture his/her preferences. Our institution is only one of a few in the US where two neurosurgeons perform DBS surgeries. Based on informal conversations with these surgeons, we started our study with the belief that different sets of weights would be necessary. Experiment 1 shows that our training mechanism, albeit heuristic, can capture surgeon preferences. The results we obtained with experiment 2 were unexpected and suggest that, in fact, the same set of weights could be used to compute trajectories for both surgeons at our institution. Experiment 3 shows that the clinical acceptance of automatic trajectories is comparable to the acceptance of trajectories manually selected by the other surgeon at our institution. Results obtained in experiment 4 show that when presented with only one trajectory and asked to rate it as acceptable or not as is the case in our pseudo-prospective study, surgeons are more selective than when comparing two trajectories without knowing their provenance as is the case in experiments 1 and 2. Indeed, most of the cases rejected pseudo-prospectively in experiment 4 were previously rated acceptable or better. The main difference between the pseudo-prospective and retrospective experiments is that in the former the surgeon knows unequivocally that the trajectory is computed. In the latter the trajectories' provenance is unknown. Even though the reason for this difference could not be elucidated with our experiments, the fact that the surgeons were stricter in the pseudo-prospective situation only reinforces the value of the results we have obtained in experiment 5. This experiment is the largest we know of and the only one that has been conducted with neurosurgeons expert in DBS

procedures at three different institutions. One surgeon has 17 years of experience with DBS procedures and has implanted more than 600 leads, one has 16 years of experience and has implanted more than 550 leads, one has 8 years of experience and has implanted more than 380 leads, and one has 4 years of experience and has implanted more than 122 leads. The results we have obtained with three of these surgeons and with 60 trajectories suggest that using a single set of weights may be adequate for multiple surgeons. The results also show that automatically computed trajectories were accepted 95% of the time by a majority of the neurosurgeons in our group and 76.67% unanimously. Acceptance rates by individual surgeons for these trajectories ranged from 83.33% to 95%, which indicate that computer-assisted assistance may be a valuable addition to DBS planning software. In our experience, eliciting priorities or even preferences from surgeons is not a straightforward task. For instance, one of the two surgeons at our institution mentioned that he likes to select trajectories that intersect the lateral edge of the thalamus (*Rule* 6). After training, the weight estimated for this rule for surgeon A was small. It ended up not being used in our pseudo-prospective study because surgeon B's weights were used and he did not utilize this rule. As shown in our experiments, surgeon A ended up preferring the weights selected for surgeon B, suggesting, as discussed earlier, that surgeon A was not as critical as surgeon B during the training phase, thus further illustrating the difficulty of capturing surgeon preferences. This was also reported by Essert *et al*. [8] and Bériault *et al.* [9]. If larger studies performed at more sites confirm the fact that the same weights can be used across institutions, it may help standardize the procedure and assist surgeons who do not have years of experience with DBS procedures. Given enough training trajectories, an alternative would be to learn surgeon preferences algorithmically.

Constraints reported in the literature but not included in our model, such as minimizing

the path length as done by Essert *et al.* [8] or the overlap with caudate as done by Bériault *et al.* [9], were not listed as reasons to reject a single trajectory. However, a recent study by Witt *et al.* [21] reports that trajectories that intersect with caudate nuclei may increase the risk of a decline in global cognition and memory performance; Benabid *et al.* [3] also suggest selecting trajectories that do not intersect the caudate. Other sites may thus have different preferences and several cost functions may have to be designed to reflect this variability, further suggesting the need for larger studies. We also note that our study focuses on STN targeting. We are now beginning to capture rules that are used for other targets such as ventrointermediate nucleus and globus pallidus interna. We will follow a procedure similar to the one described herein to determine whether automatic trajectory planning would also be useful for these targets.

As shown in Figure 2.10, the main reason for trajectory rejection is the spatial proximity to the sulci. A close inspection of the rejected trajectories shows that these penetrated the brain at the top of a gyrus, as desired, but were too close to the bottom of a sulcus. Our current approach only detects an intersection with the cortical surface and tries to determine if this intersection happens in a sulcal or gyral area, but does not penalize trajectories that are close to a sulcus once they have penetrated a gyrus. We are currently addressing this issue by localizing sulci and computing distance maps from these. We will then modify our sulcus-related cost term accordingly. In the current study, we decouple the segmentation and the path computation components by validating the segmentation visually. About 20% of the volumes required some minor editing, e.g., displacement of the ventricular boundary or enlargement of ventricular veins. Our vessel segmentation method is simple but we have compared it to the technique proposed by Frangi *et al.* [22] and have obtained similar results. When available, additional MR volumes acquired with sequences such as susceptibility weighted imaging or time-of-flight sequences as

proposed by Bériault *et al*. [9] would facilitate the segmentation of the vasculature.

Finally, integrating our solution in our clinical solution (CRAVE) for a full prospective study still requires a few improvements. Here, the target point was known and fixed. In the system in place at Vanderbilt, targets are predicted automatically and optimal trajectories can be computed for these. But, when the plan is finalized by the surgeons, the target position may be modified, which may also invalidate the pre-computed trajectory and necessitate re-computation. We are currently investigating the sensitivity of the solution to the position of target point, which is rarely moved by more than 1.5 mm. If it is very sensitive to the target point, clinical implementation may require pre-computing a series of solutions or constraining the search space to produce new solutions in clinically acceptable time.

Ultimately, computer assistance should help the surgical team in selecting a safe trajectory. Pre-computing an optimal solution and presenting it to the end user focuses him/her on what the system determines to be the safest entry point but the system also needs to permit easy interactive validation. Providing visual feedback in the form of segmented structures, risk values, distance to critical structures, or color-coded maps as others have done [4-10] and as shown in Figure 2.3 and 2.4 would also assist the surgeons in selecting an alternative entry point that may not be in the immediate vicinity of the suggested one should it be rejected.

As implemented, extracting the cortical surface takes about 2 hours to compute, segmenting the other structures 10 minutes, pre-computing the sulci cost for every entry point in the search region 10 minutes, and computing an optimal trajectory once all the pre-computations have been done a few seconds.

References

[1] M. L. Kringelbach, N. Jenkinson, S. L. Owen, and T. Z. Aziz, "Translational principles of deep brain stimulation," *Nat Rev Neurosci*, vol. 8, no. 8, pp. 623-635, 2007.

[2] R. Bakay and A. Smith, "Deep brain stimulation: Complications and attempts at avoiding them," *Open Neurosurg J*, vol. 4, pp. 42-52, 2011.

[3] A. L. Benabid, S. Chabardes, J. Mitrofanis, and P. Pollak, "Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease," *Lancet Neurol*, vol. 8, no. 1, pp. 67-81, 2009.

[4] M. Vaillant, C. Davatzikos, R. H. Taylor, and R. N. Bryan, "A path-planning algorithm for image-guided neurosurgery," in *Proc. CVRMed-MRCAS*, 1997, pp. 467-476.

[5] E. J. L. Brunenberg, A. Vilanova, V. Visser-Vandewalle, Y. Temel. L. Ackermans, B. Platel, and B. M. ter Haar Romeny, "Automatic trajectory planning for deep brain stimulation: a feasibility study," in *Proc. MICCAI*, 2007, pp. 584–592.

[6] N. V. Navkar, N. V. Tsekos, J. R. Stafford, J. S. Weinberg, and Z. Deng, "Visualization and planning of neurosurgical interventions with straight access," in *Proc. IPCAI*, 2010, pp. 1-11.

[7] R. R. Shamir, L. Joskowicz, I. Tamir, E. Dabool, L. Pertman, A. Ben-Ami, and Y. Shoshan, "Reduced risk trajectory planning in image-guided keyhole neurosurgery," *Med Phys*, vol. 39, no. 5, pp. 2885–2895, 2012.

[8] C. Essert, C. Haegelen, F. Lalys, A. Abadie, and P. Jannin, "Automatic computation of electrode trajectories for deep brain stimulation: a hybrid symbolic and numerical approach," *Int J Comput Assist Radiol Surg*, vol. 7, no. 4, pp. 517–532, 2012.

[9] S. Bériault, F. A. Subaie, D. L. Collins, A. F. Sadikot, and G. B. Pike. “A multi-modal approach to computer-assisted deep brain stimulation trajectory planning,” *Int J Comput Assist Radiol Surg*, vol. 7, no. 5, pp. 687–704, 2012.

[10] S. Bériault, S. Drouin, A. F. Sadikot, Y. Xiao, D. L. Collins, and G. B. Pike, "A prospective evaluation of computer-assisted deep brain stimulation trajectory planning,” in *Proc. CLIP*, 2013, pp. 42-49.

[11] Y. Liu, B. M. Dawant, S. Pallavaram, J. S. Neimat, P. E. Konrad, P.-F. D’Haese, R. D. Datteri, B. A. Landman, and J. H. Noble, “A surgeon specific automatic path planning algorithm for deep brain stimulation,” in *Proc. SPIE Med. Imaging*, 2012, pp. 83161D-1–83161D-10.

[12] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Trans Med Imaging*, vol. 16, no. 2, pp. 187-198, 1997.

[13] P. Viola and William M. Wells III, “Alignment by maximization of mutual information,” *Int J Comput Vis*, vol. 24, no. 2, pp. 137-154, 1997.

[14] W. E. Lorensen and H. E. Cline, “Marching cubes: A high resolution 3D surface construction algorithm,” in *Proc. SIGGRAPH*, 1987, pp. 163-169.

[15] S. K. Warfield, K. H. Zou, and W. M. Wells, “Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation,” *IEEE Trans Med Imaging*, vol. 23, no. 7, pp. 903-921, 2004

[16] J. A. Sethian, *Level Set Methods and Fast Marching Methods*, 2nd ed. Cambridge, MA: Cambridge Univ. Press, 1999.

[17] J. H. Noble, O. Majdani, R. F. Labadie, B. Dawant, and J. M. Fitzpatrick, “Automatic determination of optimal linear drilling trajectories for cochlear access accounting for drill-positioning error,” *Int J Med Robot Comp*, vol. 6, no. 3, pp. 281-290, 2010.

[18] G. K. Rohde, A. Aldroubi, and B. M. Dawant, “The adaptive bases algorithm for intensity-based nonrigid image registration,” *IEEE Trans Med Imaging*, vol. 22, no. 11, pp. 1470-1479, 2003.

[19] X. Han, D. L. Pham, D. Tosun, M. E. Rettmann, C. Xu, and J. L. Prince “CRUISE: cortical reconstruction using implicit surface evolution,” *Neuroimage*, vol. 23, no. 3, pp. 997-1012, 2004.

[20] P.-F. D’Haese, S. Pallavaram, R. Li, M. S. Remple, C. Kao, J. S. Neimat, P. E. Konrad, and B. M. Dawant, “CranialVault and its CRAVE tools: a clinical computer assistance system for deep brain stimulation therapy,” *Med Image Anal*, vol. 16, no. 3, pp. 744-753, 2012.

[21] K. Witt, O. Granert, C. Daniels, J. Volkmann, D. Falk, T. van Dimeren, and G. Deuschl, “Relation of lead trajectory and electrode position to neuropsychological outcomes of subthalamic neurostimulation in Parkinson’s disease: results from a randomized trial,” *Brain*, vol. 136, no. 7, pp. 2109-2119, 2013.

[22] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Proc. MICCAI*, 1998, pp. 130-137.

CHAPTER III

AUTOMATIC LOCALIZATION OF THE ANTERIOR COMMISSURE, POSTERIOR COMMISSURE, AND MIDSAGITTAL PLANE IN MRI SCANS USING REGRESSION FORESTS[2]

Yuan Liu and Benoit M. Dawant

Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37232 USA

[2] [This chapter has been published in *IEEE Journal of Biomedical and Health Informatics*, Volume 19, Issue 4, pp. 1362-1374, July 2015.]

Abstract

Localizing the anterior and posterior commissures (AC/PC) and the midsagittal plane (MSP) is crucial in stereotactic and functional neurosurgery, human brain mapping, and medical image processing. We present a learning-based method for automatic and efficient localization of these landmarks and the plane using regression forests. Given a point in an image, we first extract a set of multi-scale long-range contextual features. We then build random forests models to learn a nonlinear relationship between these features and the probability of the point being a landmark or in the plane. Three-stage coarse-to-fine models are trained for the AC, PC, and MSP separately using down-sampled by 4, down-sampled by 2, and the original images. Localization is performed hierarchically, starting with a rough estimation that is progressively refined. We evaluate our method using a leave-one-out approach with 100 clinical T1-weighted images and compare it to state-of-the-art methods including an atlas-based approach with six nonrigid registration algorithms and a model-based approach for the AC and PC, and a global symmetry-based approach for the MSP. Our method results in an overall error of 0.55 ± 0.30 mm for AC, 0.56 ± 0.28 mm for PC, $1.08^{\circ} \pm 0.66^{\circ}$ in the plane's normal direction and 1.22 ± 0.73 voxels in average distance for MSP; it performs significantly better than four registration algorithms and the model-based method for AC and PC, and the global symmetry-based method for MSP. We also evaluate the sensitivity of our method to image quality and parameter values. We show that it is robust to asymmetry, noise, and rotation. Computation time is 25 seconds.

3.1. Introduction

The anterior commissure (AC) and the posterior commissure (PC) are the two points with the shortest intraventricular distance between the commissures, according to the standard convention

of the Schaltenbrand-Wahren atlas [1]. They are important landmarks located in the midsagittal plane (MSP), a geometric plane that separates the two hemispheres of the cerebrum by the inter-hemispheric fissure (IF). The AC and PC, together with the MSP, define a standardized coordinate system widely used by major stereotactic brain atlases such as the Schaltenbrand-Wahren atlas [1] and the Talairach and Tournoux atlas [2]. Establishing this reference system is pivotal for stereotactic and functional neurosurgery, human brain mapping, and medical image processing [3-5]. For example, in deep brain stimulation (DBS) procedures, target locations are estimated by their relative positions in this standardized system [3]. Identification of the AC, PC, and MSP could also facilitate the estimation of an initial intra- or inter-subject affine transformation to reduce the degrees of freedom in nonrigid transformations used to register two image volumes [6]. Yet another example is the quantification of the structural and radiometric asymmetry of the brain made possible by the localization of the MSP. This can be used to detect brain pathologies such as tumors that cause severe asymmetry between the two hemispheres [7].

In most current neuroimaging applications, the AC, PC, and MSP are selected manually in the magnetic resonance image (MRI) scans by experts. This requires expertise and suffers from inter-expert variability, which can have a substantial effect on targeting in image guided neurosurgery [8]. Manual intervention also takes time and prevents the automated use of such information by other image processing techniques such as registration. Over the years, several approaches have thus been proposed to automatically localize the AC and PC [6], [9-14] as well as the MSP [15-24] on 3D MRI scans.

For the AC and PC, these algorithms rely on the successful segmentation of surrounding structures, the localization of other anatomical landmarks, or image registration. For example, in [6][9-11], the corpus callosum was used to initialize the AC and PC positions. Ardekani *et al.*

[12] achieved the initialization by identifying the MSP and a landmark on the midbrain-pons junctions. Han *et al.* [6] and Verard *et al.* [9] also relied on edge detection. In [13-14], atlas-based nonrigid registration was performed to transfer the AC and PC positions from atlases onto subjects. However, brain segmentation, landmark identification, edge detection, and nonrigid registration algorithms may fail due to large anatomical variations or image contamination by noise or partial volume effect, leading to the failure of the AC/PC detection. In addition, some of these methods require long runtime, especially for registration based methods.

For the MSP, most existing methods can be categorized into two types: 1) methods maximizing a global symmetry score, 2) methods detecting the IF. The first type of approaches assumes global bilateral symmetry and maximizes a similarity measure between the original brain scan and its reflected version [15-18]. However, there is no perfect bilateral symmetry in the human brain, not only for pathological cases but even for normal cases. As shown in Figure 3.1 for a control subject, an effect known as brain torque occurs when the left occipital lobe or the right frontal lobe is larger than its counterpart in the other hemisphere [25]. Hence these methods may suffer from sensitivity to brain asymmetry and also often from high computational cost, while they may generalize well to other image modalities. On the other hand, approaches of the second type identify the IF from its intensity and textural features or by locally optimizing a symmetry measure, as local symmetry could be assumed in the vicinity of the IF region. The MSP is then determined by fitting a plane to those detected points or line segments [19-23]. These methods are generally more robust to abnormalities but more sensitive to outliers in the set of feature points. A robust outlier removal method is usually required to achieve the desired accuracy.

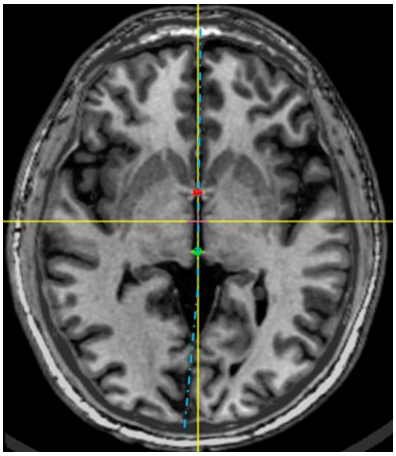

Figure 3.1. An example of the brain torque effect. The MSP represented as the vertical yellow axis deviates in the posterior region from the blue dotted curve which separates the hemispheres symmetrically in this slice.

Recently, learning-based methods using random forests have gained popularity for landmark and plane detection. Random forests are an ensemble supervised learning technique for classification or regression. In this approach a multitude of decision trees are constructed by evaluating a random subset of features at each node to split the data. The output of these trees is then aggregated to produce a final prediction [26]. In [27], Dabbah *et al.* used random forests as a classifier to localize anatomical landmarks in CT. Hough forests, which combine random forests with generalized Hough transform, are used to detect points to drive an active shape model on 2D radiographs [28], to find a rough position for the center of vertebrae in MR images [29], and most recently to localize the parasagittal plane in ultrasound images [30]. Schwing *et al.* [24] proposed to use adaptive random forests to jointly identify five distinct landmarks in the MSP in T1-weighted (T1-w) images and estimate the plane via a least squares fit of these landmarks.

We have previously proposed a learning-based framework using regression forests to detect the AC and PC [31]. Here, we extend our previous work and augment it by also localizing the MSP. Since the AC, PC, and MSP have different local appearances from other points in the

image, we hypothesize that a nonlinear regression can be used to estimate the relationship between the local appearance of a point and its probability to be the AC, the PC, or in the MSP. Compared to exiting techniques, our method is conceptually the most similar to the approach of Schwing *et al.* [24] in that a learning framework is used in both to detect the plane. We do, however, regress the distance from a point to the plane directly as opposed to its distance to selected landmarks as done in [24] because those landmarks may not be a perfect indicator of the plane due to the brain asymmetry.

The algorithm we propose is fast, accurate, and robust. It does not rely on any preprocessing of the images such as edge enhancement, nor does it require any segmentation or registration. Instead, we extract multi-scale contextual features for points in a set of training images and build random forests regression models to learn the probability for each sample to be the AC, the PC, or in the MSP. We employ three-stage coarse-to-fine models, with the first one operating on a down-sampled image to roughly localize the landmark or the plane and the second and third models to fine-tune the position. We evaluate our algorithm in a leave-one-out fashion using a large clinical dataset of 100 subjects. We also compare our method to state-of-the-art methods including an atlas-based approach with six well-established nonrigid registration algorithms and a publicly available implementation of a model-based approach for the AC and PC, as well as a publicly available implementation of a global symmetry-based approach for the MSP. We further test the sensitivity of our algorithm to anatomic abnormality, image quality, and parameter values.

3.2. Methods

3.2.1. Data

We select 100 subjects from a data repository we have created over a decade for DBS surgeries [32]. All images in our data set are T1-w sagittal MR volumes with approximately 256×256×170 voxels and 1 mm in each direction, acquired with the SENSE parallel imaging technique (T1W/3D/TFE) on a 3 Tesla Phillips scanner (TR = 7.92 ms, TE = 3.65 ms). These images have similar pose with small differences in head orientation and position, and also similar field of view (FOV), i.e., they cover the entire head. All images have been acquired as part of the normal delivery of care and every subject was consented to participate in this study.

For the AC and PC, two raters manually identified these points for each subject. These two raters followed the same protocol to select the AC and PC and were given sufficient time for accurate localization. The inter-rater variability is 0.57±0.47 mm for the AC and 0.57±0.37 mm for the PC. Gold standard AC and PC are computed as an average of the selections by the two raters.

For the MSP, one rater manually selected the plane for each subject. Given the gold standard AC and PC, the MSP could be defined using any other point on the IF. However, cerebral atrophy that affects some DBS patients results in a widening of the IF. This makes the point selection on the IF ambiguous. In order to uniquely define the MSP, we follow the approach used clinically by an experienced neurosurgeon which relies on the falx cerebri, as illustrated in Figure 3.2. The falx cerebri is a sickle-shaped fold of dura mater that descends vertically in the IF [33]. It is usually visible in CT but not in T1-w images. Hence we used the CT volumes of the same patients and rigidly registered them to their corresponding T1-w volumes. After the registration, the CT image and the gold standard AC and PC were loaded into a visualization software. A random point on the falx cerebri was selected first to establish an

initial AC-PC coordinate system. The origin of this coordinate system is defined as the midpoint between the AC and the PC, with the x-axis perpendicular to the MSP pointing to the right, the y-axis pointing from the PC to the AC, and the z-axis pointing superiorly in the MSP. The CT image was then resampled in this coordinate system in which the xz planes correspond to the coronal slices. The MSP was then refined using coronal slices by varying the orientation of the plane to align it with the falx cerebri as accurately as possible. As the brain torque effect mostly causes the MSP to curve in the anterior and posterior regions of the brain, we used the midbrain region to define it, i.e., we used coronal slices that are anterior to the PC and posterior to the AC. After manual adjustment, the AC-PC coordinate system was updated and the CT image resampled in the new system for visual check in axial, coronal, and sagittal views. This process was repeated until the plane was visually deemed to be satisfactory.

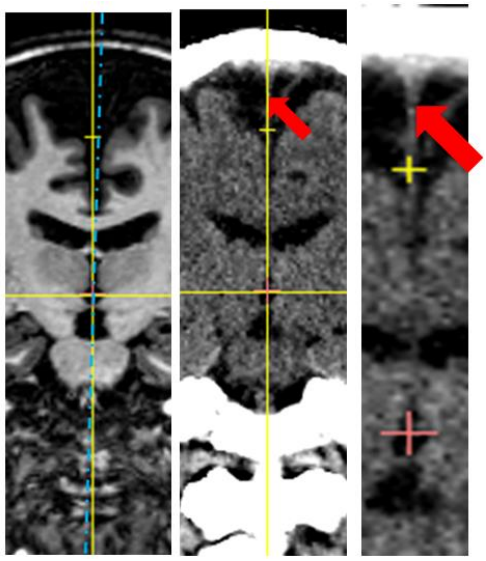

Figure 3.2. Illustration of the MSP selection. The vertical yellow axis is the selection based on the falx cerebri, shown as the bright line on the right pointed by the red arrow, and the blue dotted line is one possible plane when selecting the MSP in the T1-w image.

The gold standard AC and PC points together with the manual selection of the MSP are used as ground truth for the training and the evaluation of our learning-based method.

3.2.2. Problem Formulation

We use a voxel-level training solution based on regression forests [6]. For each voxel, we extract a set of features that describes textural context variation at different scales, as proposed by Pauly *et al.* [34]. This is realized by applying a random displacement to a voxel x, calculating the mean intensities of a 3D cuboidal region R_x^s centered on x and of a similar region $Q_x^{s,m}$ of the same size but centered on the displaced voxel, and subtracting these two:

$$f^m = \frac{1}{|R_x^s|}\left(\sum_{x' \in Q_x^{s,m}} I(x') - \sum_{x' \in R_x^s} I(x')\right) \tag{3.1}$$

where I is the intensity, and s is the current scale, i.e., a particular size of the cuboidal region. Four scales are used and they correspond to window sizes of 4, 8, 16, and 32. This process is repeated $M = 2000$ times to obtain the feature set $\{f^m\}_{m=1}^M$.

Each voxel is associated with a probability p to be the AC, the PC, or in the MSP that a model is trained to detect. This probability follows a truncated Gaussian distribution based on its Euclidean distance d to the ground truth AC, PC, or MSP:

$$p = \begin{cases} e^{-\frac{d^2}{2\sigma^2}} & d > d_{th} \\ 0 & d \le d_{th} \end{cases} \tag{3.2}$$

where σ is the standard deviation of the Gaussian function. We truncate this function at $p = 0.1$ to speed up the training process.

Given a number of training pairs $\{\vec{f_n}, p_n\}_{n=1}^N$, the random forests aim to learn a nonlinear mapping from the feature space $\{\vec{f}\}$ to the probability space $\{p\}$. Hence the AC, PC, and MSP

localization problem can now be formulated as first finding a set of voxels in the images that have a high probability to be the AC, PC, or in the MSP and then use these points to localize the landmarks or the plane.

3.2.3. Regression Forests

We use 20 regression trees to construct the forest. For each tree, a bootstrap of two thirds of the training samples is randomly selected and fed to the root node of the tree. Given the training samples $\{\overrightarrow{f_n}, p_n\}_{n=1}^{N'}$ at a particular node, we seek to select a feature f^m and a threshold t to best split the data in order to minimize the mean squared error MSE:

$$t, m = \arg\min_{t,m} \left(MSE(\{p_n : f_n^m < t\}) + MSE(\{p_n : f_n^m \geq t\})\right) \tag{3.3}$$

A subset of 500 features is randomly selected to estimate the splitting threshold. A tree stops growing if the number of samples arriving at leaf nodes is smaller than 5 or if the best split threshold cannot be found.

Each leaf of the regression trees stores the mean probability of all samples arriving at that node to be a point of interest and this is used as a predictor. When a test sample comes, each tree contributes to a prediction. The mean and the variance of these predictions across trees are calculated and outputted for this test sample.

3.2.4. Training Phase

We train separate models for the AC, PC, and MSP. For each, we build three stage coarse-to-fine models, one on down-sampled by 4 images, one on down-sampled by 2 images, and one on full resolution images. This results in nine models to train for one training dataset.

Sampling Strategy

Since all images in our dataset have similar pose and FOV, it is sufficient to search for the landmarks or the plane in a region of interest instead of searching the entire image. Hence, when training the model, for each image, we evaluate a set of points within a region of interest.

For the AC and PC, this region of interest is a cube centered on the ground truth points at each resolution level. It is a $15 \times 15 \times 15$ voxel^3 cube that covers a $60 \times 60 \times 60$ mm^3 volume at the coarsest level, a size that we have found large enough to account for the variations in the AC and PC positions across all images in the dataset. For each training subject, we use all the voxels in these cubes to generate the training samples.

For the MSP, at the coarse level, we follow the same strategy we use for the AC and PC to define a pseudo landmark in the MSP and define a cube centered on this point to be the region of interest. This point is selected to be on the z-axis, i.e., in the MSP plane, +50 mm away from the origin in the AC-PC coordinate system established by the ground truth AC, PC, and MSP. We refer to this point as the mid-plane point (MP) and calculate the MP for each training subject. We use all the voxels in the $15 \times 15 \times 15$ voxel^3 cube centered on the MP as training samples as we have done for the AC and PC. For the next two resolution levels, as the MSP can already be roughly localized at the coarsest level, we only need to sample points that are spatially close to the MSP to further distinguish them from the true MSP. Hence we define the region of interest to be a rectangular cuboid encompassing the plane and aligned with the axes of the AC-PC coordinate system. At the down-sampled by 2 level, coordinates of the lower left corner and upper right corner are (-15 mm, -15 mm, -30 mm) and (+15 mm, +15 mm, +90 mm) respectively. At the full resolution level, we narrow the region of interest to [-7 mm, +7 mm] in the x, i.e., lateral direction, for further refinement. Voxels in the original image space are transformed into the AC-PC space first to check whether they fall into the region of interest, and those that satisfy

this condition are considered as candidates. This leads to a large number of candidate points for training that cannot fit into the main memory. We address this issue by randomly sampling a fixed number of candidate points for each training subject.

An example of the training regions from which samples are drawn at the different resolution levels is shown in Figure 3.3, with the top row illustrating the AC regions and the bottom row illustrating the MSP regions.

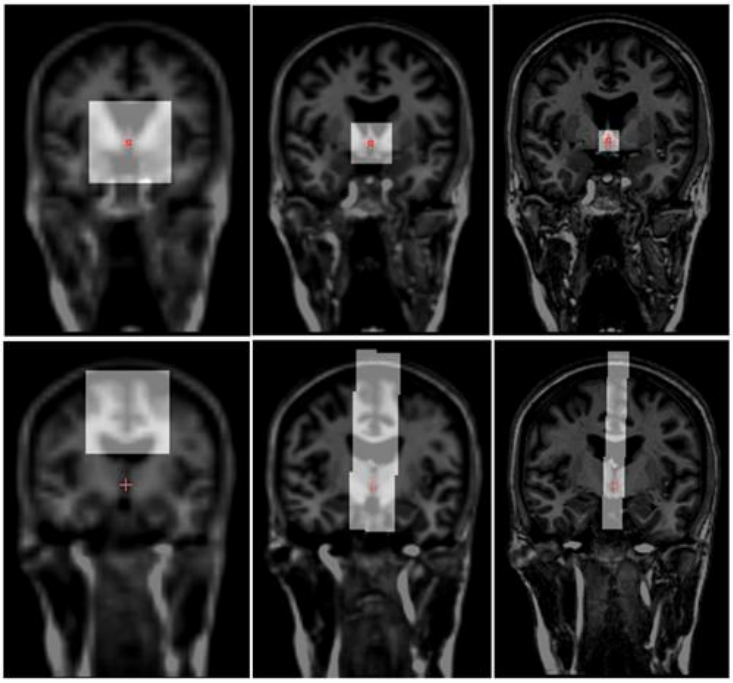

Figure 3.3. Sampling regions for the AC (top row) and the MSP (bottom row) of a training subject at down-sampled by 4 (left), down-sampled by 2 (middle), and full resolution level (right), each overlaid on the corresponding image.

3.2.5. Testing Phase

Given a test image, following the hierarchical approach described earlier, we first down-sample the image by 4 and start testing using the models built at this resolution level. At each level, we sequentially test for the AC, PC, and MSP. We initialize the search center from this model by

averaging the ground truth AC, PC, and MP of all training subjects, and search within a $21 \times 21 \times 21$ voxel^3 window on a regular grid. A response map is generated for each landmark, displaying the mean probabilities $\bar{p}$ of the voxels to be the trained landmark. For the AC and PC, the voxel associated with the highest probability is then used as the search center for the next level. For the MSP, a plane is estimated from the response map for the MP together with the AC and PC detected at this level, and this estimation is used to define the search region for the next level. We will explain in detail how to estimate the plane later. We continue this testing process for the next two resolution levels. For those two levels, the testing samples for the AC and PC are the points within a $21 \times 21 \times 21$ voxel^3 volume with its search center estimated from the previous level. For the MSP, the AC-PC coordinates of all voxels are first calculated based on the AC, PC, and MSP estimated from the previous level. The testing samples are those whose AC-PC coordinates are within [-15 mm, +15 mm] in the x direction, [-15 mm, +15 mm] in the y direction, and [-30 mm, +90 mm] in the z direction at the down-sampled by 2 level, and [-7 mm, +7 mm] in the x direction, [-15 mm, +15 mm] in the y direction, and [-30 mm, 90 mm] in the z direction at the full resolution level. An example of the response maps at different levels for one testing subject is illustrated in Figure 3.4, with the top row illustrating the AC response maps and the bottom row illustrating the MSP response maps.

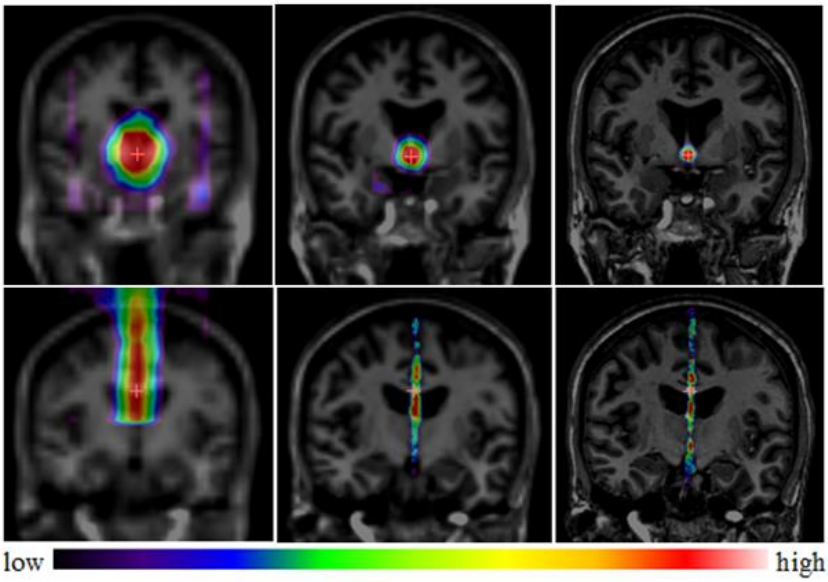

Figure 3.4. Response maps for the AC (top row) and the MSP (bottom row) of a testing subject at down-sampled by 4 (left), down-sampled by 2 (middle), and full resolution level (right), each overlaid on the corresponding image.

Mean-shift Refinement for the AC/PC

When reaching the full resolution level, the final prediction for the AC and PC is made via weighted mean shift, a technique that is used to localize the maximum of a density function given discrete data sampled from that function [35]. Starting with an initial estimate $x^{(0)}$, mean shift iteratively re-estimates the mean by weighting each point in its neighborhood, with the value of the weight determined by its distance to the current estimate and the probability of this point to be the trained landmark:

$$x^{(t+1)} = \frac{\sum_{x_i{}^{(t)} \in N(x^{(t)})} K\left(x_i{}^{(t)} - x^{(t)}\right) \cdot \overline{p_i{}^{(t)}} \cdot x_i{}^{(t)}}{\sum_{x_i{}^{(t)} \in N(x^{(t)})} K\left(x_i{}^{(t)} - x^{(t)}\right) \cdot \overline{p_i{}^{(t)}}} \tag{3.4}$$

where $x^{(t)}$ is the current estimate, $x_i{}^{(t)}$ is a point in the neighborhood $N(x^{(t)})$ whose distance to $x_i{}^{(t)}$ is smaller than 6, $K\left(x_i{}^{(t)} - x^{(t)}\right) = e^{-k\left\|x_i{}^{(t)} - x^{(t)}\right\|^2}$ is the Gaussian kernel function, and

$\overline{p_i^{(t)}}$ is the value of $x_i^{(t)}$ in the response map which corresponds to the probability averaged across all trees.

We choose the initial estimate $x^{(0)}$ to be the position of the testing sample with the maximum probability in the response map and iteratively update $x^{(t)}$ until it converges. The output of this estimate is the final prediction for the AC/PC.

Weighted Least Squares Fitting for the MSP

We represent the MSP as:

$$ax + by + cz + d = 0, s.t. \|[a, b, c]\| = 1 \tag{3.5}$$

where a, b, c, d are the parameters that uniquely define the plane.

At each testing stage, we need to estimate the MSP using its response map along with the current AC and PC estimates. This is done by selecting a set of points that have a high probability $\bar{p}$ to be in the MSP and fitting a plane to those points. To perform a robust linear regression, we compute a weighted least squares solution to account for the degree of uncertainty per point:

$$\min_{a,b,c,d} \sum_{i=1}^{N'} (ax_i + by_i + cz_i + d)^2 \cdot w_i \,, s.t. \|[a, b, c]\| = 1 \tag{3.6}$$

where N' is the number of candidates, and w_i is the associated weight for sample (x_i, y_i, z_i).

For each candidate, we define its weight to be the square of the mean of the predictions divided by the variance of the predictions across all trees:

$$w_i = \overline{p_i}^2 / \mathrm{Var}(\overrightarrow{p_i}) \tag{3.7}$$

where $\overline{p_i}$ indicates how close this point is to the MSP as predicted by the model, and $\mathrm{Var}(\overrightarrow{p_i})$ indicates the model's confidence about $\overline{p_i}$.

The final MSP prediction is the weighted least squares estimate at the full resolution level.

3.2.6. Comparison to Other Methods

For the AC and PC, we compare our results with those obtained by atlas-based registration, including affine only, and affine + nonrigid registration, a technique routinely used for automatic identification in DBS procedures [36]. We choose one atlas used by Pallavaram *et al.* [36] to be the reference and project its AC and PC points onto the 100 subjects through registration. An affine transformation is estimated first by an intensity-based technique that uses Mutual Information [37-38] as its similarity measure. The results are visually checked and manually corrected if a failure is observed. Then nonrigid registration is performed with a series of well-established algorithms, including the Adaptive Basis Algorithm (ABA), a variation of ABA that is tuned for deep brain structures referred to as the Adaptive Basis Algorithm with bounding box (LABA) [39], Diffeomorphic Demons (DD) [40], Symmetric Normalization (SyN) [41], Fast Free Form Deformation (F3D) [42], and Automatic Registration Toolbox (ART) [43]. A detailed description of those algorithms can be later found in Chapter IV [44]. We also compare our method to a publicly available toolkit called YUKI which implements a recently proposed model-based approach to detect the AC and PC [12].

For the MSP, we compare our method to the same toolkit YUKI, which also implements a global symmetry-based approach to localize the plane [15].

3.3. Results

3.3.1. Leave-one-out Validation

We have conducted a leave-one-out validation, which uses 99 volumes for training and the last one for testing, and repeats this process 100 times.

Results for the AC and PC

A representative example of response maps at the full resolution level is shown in Figure 3.5, with the top row showing the response map for AC and the bottom row showing the one for PC. Each map is overlaid on top of the original images, with the cross indicating the ground truth and the white dot our prediction. As shown in this figure, the ground truth AC and PC have high probabilities and are close to our estimations (0.55 mm differences for both AC and PC).

To quantitatively evaluate the accuracy of the algorithm, we use the 3D Euclidean distance between the automatically detected landmarks and the ground truth. We refer to our algorithm as RF (Random Forests) in the following text. Figure 3.6 shows the boxplot of errors using different methods for the AC and PC. There are some outliers with errors beyond the maximum range of the y-axis (12 mm), which are not shown in the figure. This includes 2 cases using Affine, 4 cases using YUKI for the AC, and 4 cases using YUKI for the PC. We also report error statistics for the AC in Table 3.1 and PC in Table 3.2. We have excluded those above-mentioned outliers when computing mean, maximum, and standard deviations so as not to bias the comparisons. Table 3.1 and 3.2 demonstrate that our method leads to smaller mean, maximum, and standard deviation of errors for the AC and PC than the registration-based methods using Affine, ABA, DD, F3D, and ART, as well as the toolkit YUKI.

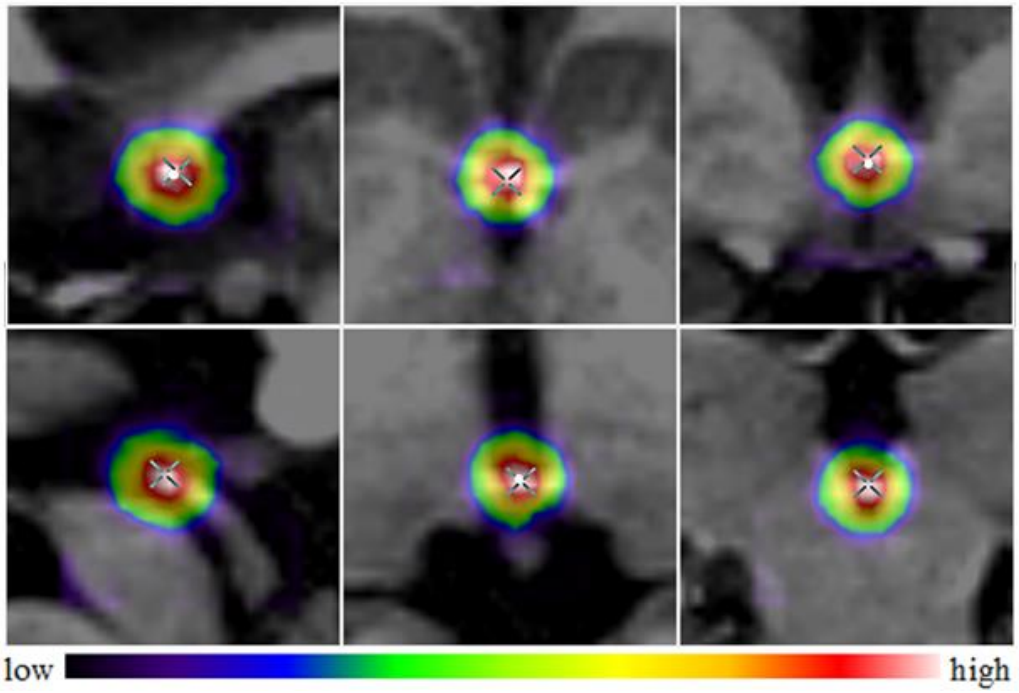

Figure 3.5. A representative example of response maps for AC (top row) and PC (bottom row) in sagittal (left), axial (middle), and coronal (right) views.

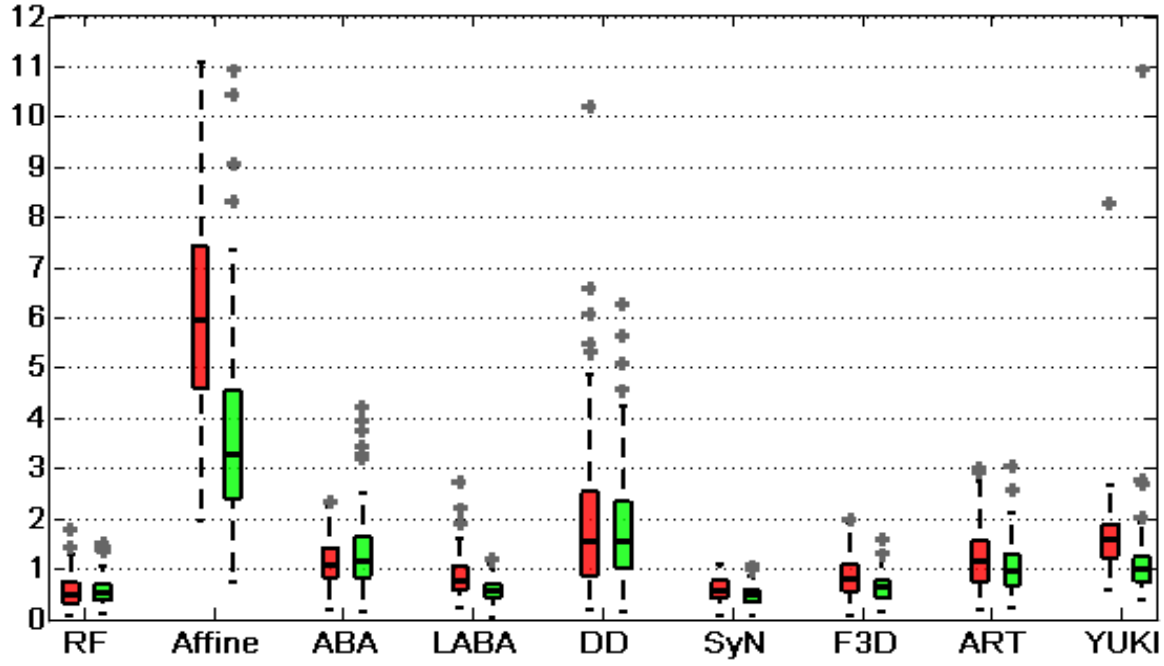

Figure 3.6. Boxplot of errors for the AC (red) and PC (green) in millimeters.

Table 3.1. Statistics of errors between the automatically detected AC using different methods and the ground truth AC positions

Error (ε): (mm) (AC)	Cases with $\boldsymbol{\varepsilon}$ <1	Cases with 1≤ $\boldsymbol{\varepsilon}$ <2	Cases with 2≤ $\boldsymbol{\varepsilon}$ <3	Cases with $\boldsymbol{\varepsilon}$ ≥3	Mean	Max.	Std.
RF	95	5	0	0	0.55	1.78	0.30
Affine	0	1	2	97	5.96	11.06	1.87
ABA	43	53	4	0	1.12	2.35	0.48
LABA	73	25	2	0	0.83	2.73	0.40
DD	29	32	22	17	1.96	10.18	1.58
SyN	94	6	0	0	0.60	1.10	0.24
F3D	67	33	0	0	0.86	1.98	0.39
ART	40	52	8	0	1.19	2.99	0.60
YUKI	17	65	13	5	1.58	8.26	0.82

Table 3.2. Statistics of errors between the automatically detected PC using different methods and the ground truth PC positions

Error (ε): (mm) (PC)	Cases with $\boldsymbol{\varepsilon}$ <1	Cases with 1≤ $\boldsymbol{\varepsilon}$ <2	Cases with 2≤ $\boldsymbol{\varepsilon}$ <3	Cases with $\boldsymbol{\varepsilon}$ ≥3	Mean	Max.	Std.
RF	93	7	0	0	0.56	1.50	0.28
Affine	1	13	28	58	3.68	10.90	1.88
ABA	38	51	5	6	1.33	4.23	0.75
LABA	93	7	0	0	0.57	1.18	0.23
DD	23	43	22	12	1.83	6.25	1.15
SyN	99	1	0	0	0.47	1.07	0.19
F3D	90	10	0	0	0.66	1.61	0.28
ART	54	43	2	1	1.02	3.04	0.49
YUKI	51	41	3	5	1.13	10.91	1.09

We also performed one-sided paired Wilcoxon signed-rank statistical tests to determine whether or not the medians of errors using our method are smaller than those using the other methods. The *p*-values are shown in Table 3.3. *P*-values smaller than the significance level 0.05 are marked in red bold. These results show that our method is significantly better than most of

the other methods. Indeed, it significantly reduces the AC and PC localization errors compared to the registration-based approaches using Affine, ABA, DD, F3D, and ART, as well as the publicly available toolkit YUKI.

Table 3.3. *P*-values of Wilcoxon tests between the errors of AC/PC detected by our method and errors of those detected by other automatic methods

P-values	Affine	ABA	LABA	DD	SyN	F3D	ART	YUI
AC	**0.00**	**0.00**	**0.00**	**0.00**	0.07	**0.00**	**0.00**	**0.00**
PC	**0.00**	**0.00**	0.23	**0.00**	1.00	**0.00**	**0.00**	**0.00**

Results for the MSP

A representative example of response maps at the full resolution level is shown in Figure 3.7, with the maps overlaid on top of the full resolution images resampled in the AC-PC coordinate system established by the ground truth AC, PC, and MSP. The vertical yellow axis represents the ground truth MSP and the red line represents the MSP detected by our algorithm. As shown in Figure 3.7, points in the plane have high probabilities and our estimation is close to the manual selection (1.10˚ difference in the normal direction).

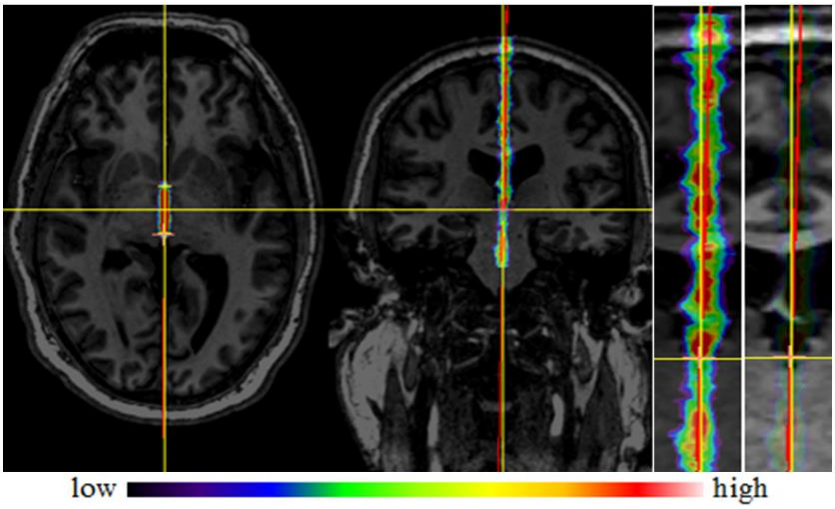

Figure 3.7. A representative example of response maps for the MSP in the axial (left), coronal (middle) and zoomed-in coronal (right) views.

The best, an average, and the worse results are shown in Figure 3.8 in the original image space. As before the red lines represent our estimations and the yellow lines the ground truth.

To quantitatively evaluate the accuracy of our algorithm, we use the angular differences in the normal direction between the automatically detected MSP and the ground truth MSP as a measure of error. We also use a metric called average z-distance proposed by Ruppert *et al.* [18]. Referring to (3.6), this metric is computed as the absolute difference in the z coordinate in voxels between the ground truth plane and the estimated plane, averaged for all possible pairs of x and y values. This measure provides a sense of the average distance between the planes. Figure 3.9 shows the boxplot of errors in the normal direction and in the average z-distance for our method and for YUKI. There are some outliers with errors beyond the maximum range of the y-axis (6˚ or 6 voxels), which are not shown in the figure. This includes 3 cases using YUKI. We also report error statistics for the normal direction in Table 3.4 and for average z-distance in Table 3.5. We have excluded those above-mentioned outliers for YUKI when computing mean, maximum, and standard deviations in order not to bias the comparisons. Table 3.4 and 3.5 demonstrate that our method leads to smaller mean, maximum, and standard deviation of errors in the normal

direction and in the average z-distance compared to the toolkit YUKI.

We also performed one-sided paired Wilcoxon signed-rank statistical tests to determine whether or not the medians of errors using our method are smaller than those using YUKI. *P*-values for both the angular errors and the average z-distance are smaller than the significance level 0.05, indicating that our method significantly outperforms YUKI.

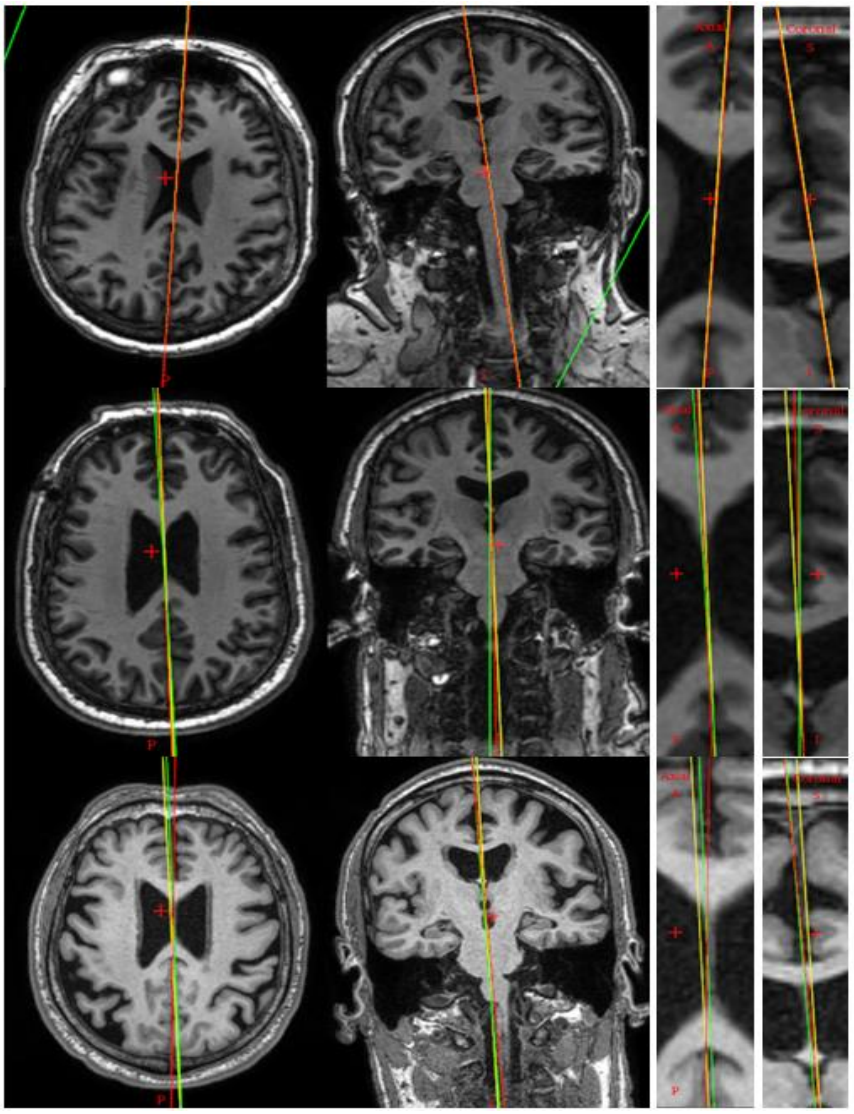

Figure 3.8. The best (top row), an average (middle row), and the worst (bottom row) MSP results using our proposed method. Ground truth MSPs are shown as yellow lines, our results as red lines, and the YUKI results as green lines.

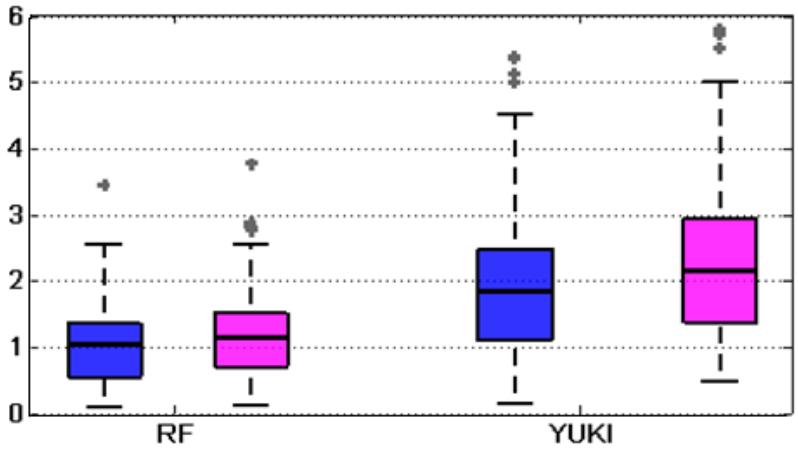

Figure 3.9. Boxplot of errors for the MSP in the normal direction in degrees (blue) and in average z-distance in voxels (magenta).

Table 3.4. Statistics of errors in the normal direction between the automatically detected MSP using different methods and the ground truth MSP

Error (ε): ($^\circ$) (MSP)	Cases with $\boldsymbol{\varepsilon}$ <1	Cases with 1≤ $\boldsymbol{\varepsilon}$ <2	Cases with 2≤ $\boldsymbol{\varepsilon}$ <3	Cases with $\boldsymbol{\varepsilon}$ ≥3	Mean	Max.	Std.
RF	49	39	11	1	1.08	3.43	0.66
YUKI	22	35	30	13	1.92	5.38	1.11

Table 3.5. Statistics of errors in average z-distance between the automatically detected MSP using different methods and the ground truth MSP

Error (ε): (voxel) (MSP)	Cases with $\boldsymbol{\varepsilon}$ <1	Cases with 1≤ $\boldsymbol{\varepsilon}$ <2	Cases with 2≤ $\boldsymbol{\varepsilon}$ <3	Cases with $\boldsymbol{\varepsilon}$ ≥3	Mean	Max.	Std.
RF	42	44	13	1	1.22	3.76	0.73
YUKI	13	30	33	24	2.25	6.12	1.20

3.3.2. Robustness Evaluation

We have conducted a series of experiments to assess the robustness of our method with regard to asymmetry of the brain, quality of the images, and rotational variations across subjects. Our experimental design is similar to those described by Liu *et al.* [16] and Hu *et al.* [23], except that we use clinical volumes instead of mirrored images. For each experiment, we randomly select 10

subjects from the dataset, generate simulated images under certain conditions, and localize target anatomies using the models previously trained for the unperturbed volume as described in section 3.3.1. We measure the localization accuracy with the metrics used in section 3.3.1, i.e., the distance error for AC and PC, the angular error and the distance error (average z-distance) for MSP. Results are averaged over the 10 subjects and shown in the following subsections.

Robustness with respect to Brain Asymmetry

We simulate two scenarios to introduce various pathologies that cause brain asymmetry. The first is to superimpose a spherical lesion with a specified position, radius, and intensity value in each test volume. The intensity value of the sphere replaces the value of the original voxels, as done in [23]. The second is to apply a local growth model with a specified seed point, radius, and growth rates to deform brain tissues. The growth model is a radially symmetric displacement field that produces deformations originating from the seed point and spreading gradually over the brain. The seed position is chosen to be away from the MSP to induce asymmetric deformation by the model. More details on this model can be found in [45]. For each test subject, we generate a set of simulated images with a range of radii for the sphere or the growth model as illustrated in Figure 3.10 and localize the AC, PC, and MSP in those images.

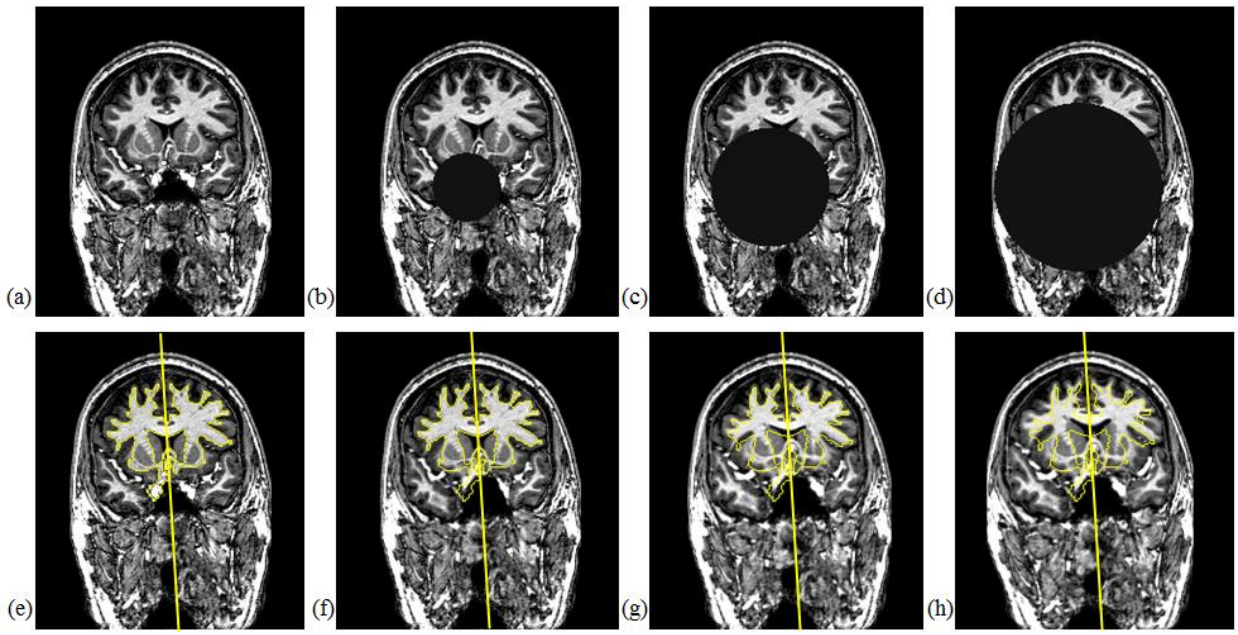

Figure 3.10. Examples of simulated brain asymmetry for a test subject. Panel (a) shows one slice of the original image, and (e) is the same slice overlaid with the ground truth MSP and segmented structures represented as yellow contours. Panels (b)-(d) show the images superimposed with spherical lesions of radius 30 mm, 50 mm, and 70 mm respectively, with all center points located at (XSize/2-20, YSize/2+20, ZSize/2-10) and intensity value = 20. Panels (f)-(h) are the images deformed by growth models of radius 30 mm, 50 mm, 70 mm respectively, with all seed points located at (XSize/2-20, YSize/2+20, ZSize/2-10).

Figure 3.11 shows the localization accuracy for the AC, PC, and MSP for different sizes of spherical lesions averaged over the 10 subjects. As shown in the figure, the AC and PC are localized well for spheres of small size; errors increase drastically when the radius is larger than 40 mm. This is because the sphere has occluded the AC/PC region. In this situation, no valid AC/PC point exists in the image and detecting those points logically fails. Localization of the MSP is less sensitive to the size of the sphere, as illustrated by a successful case in Figure 3.12 (a). An exception occurs for one subject when the radius is larger than 40 mm, as shown in Figure 3.12 (b). This is because we use the detected AC and PC to help define the search region for MSP, and the failure of AC/PC detection in this case leads to a highly skewed region with little coverage of the MSP. In practice, this scenario is unlikely to happen. If it does happen, our algorithm needs to be modified to handle failure cases for AC/PC so that these will not affect the

MSP detection.

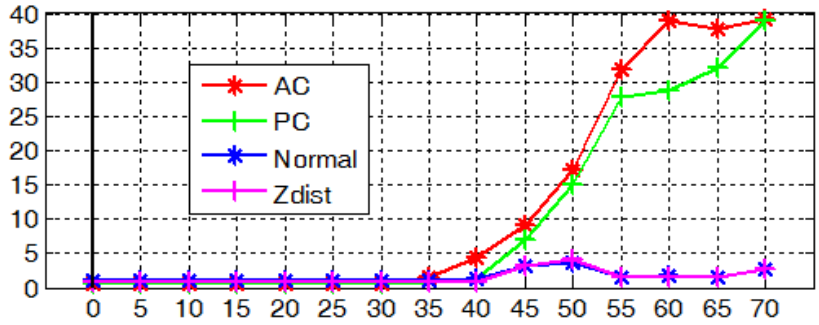

Figure 3.11. Localization errors for the AC, PC, and MSP in millimeters, degrees, or voxels with simulated spherical lesions of varying radius in millimeters.

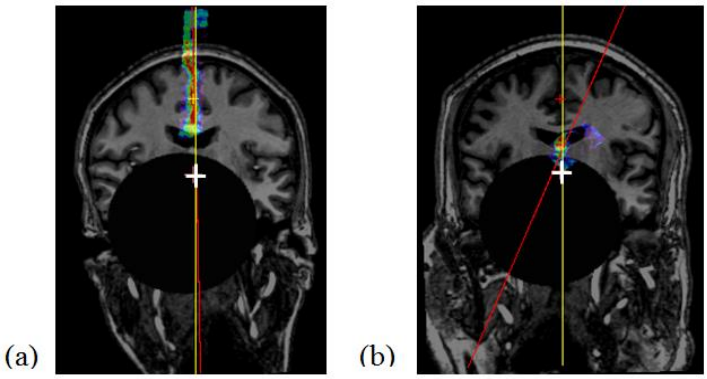

Figure 3.12. A successful (a) and a failed (b) case for MSP detection with a spherical lesion of radius = 50 mm. Both images are resampled in the AC-PC reference system with their response maps for the MSP overlaid on top. The white cross indicates the ground truth AC and PC projected onto this slice, the yellow line the ground truth MSP, and the red line the detected MSP.

Figure 3.13 shows the localization accuracy for the AC, PC, and MSP for different levels of tissue deformation averaged over the 10 subjects. As shown in this figure, the deformation we introduced only has a moderate impact on the results, with localization errors for the AC, PC, and MSP within 1.5 mm, degree, or voxel. The errors for MSP increase slightly with the amount of deformation. This is because the MSP deviates from planarity in those images; approximating it by a plane as we do in (3.5) is thus no more accurate. Figure 3.14 illustrates such example for a

test case, in which the effect of the tissue deformation is shown with a deformed grid (b) and an MSP detection result (c).

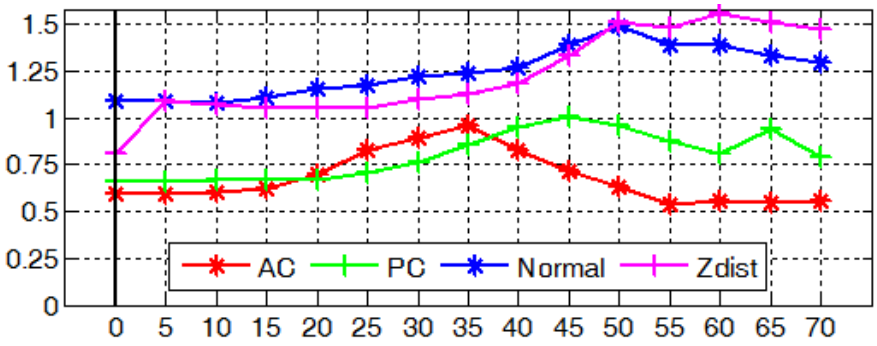

Figure 3.13. Localization errors for the AC, PC, and MSP in millimeters, degrees, or voxels with simulated tissue deformation using growth models of varying radius in millimeters.

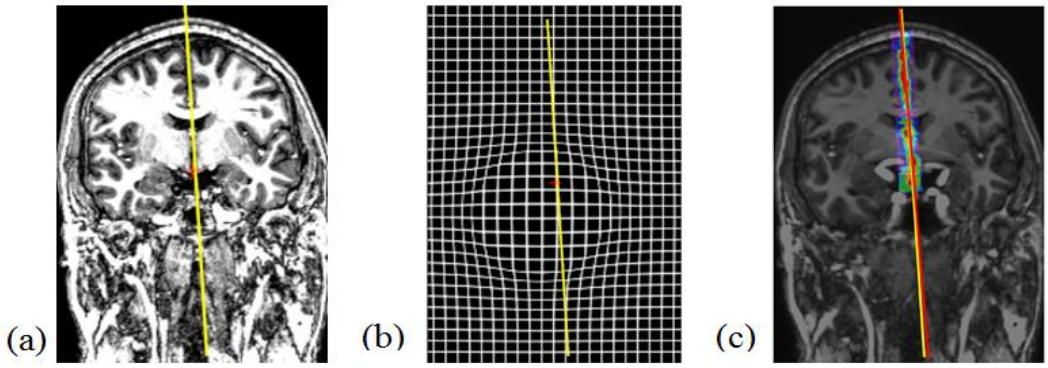

Figure 3.14. A test case with a growth model of radius = 50 mm. Panel (a) is the original image, (b) is the grid deformed by the growth model, and (c) is the deformed volume overlaid with the response map for MSP, with the yellow line representing the ground truth MSP and the red line the detected MSP.

Robustness with respect to the Noise Level

We artificially degrade each test volume by adding Gaussian white noise with zero mean and different variances. The localization accuracy of the AC, PC, and MSP for those degraded volumes averaged over 10 subjects at the same signal-to-noise ratio (SNR) level measured in decibel scale is shown in Figure 3.15. As shown in the figure, our algorithm is able to detect AC

and PC with mean error around 1 mm and extract MSP with mean angular error within 1.5˚ and mean distance error within 1.5 voxel, even when the SNR is as low as -5 dB. Results for a test case with SNR = -5 dB are shown in Figure 3.16, with 1.53 mm error for AC, 1.82 mm for PC, 1.06° angular error and 1.22 voxel distance error for MSP.

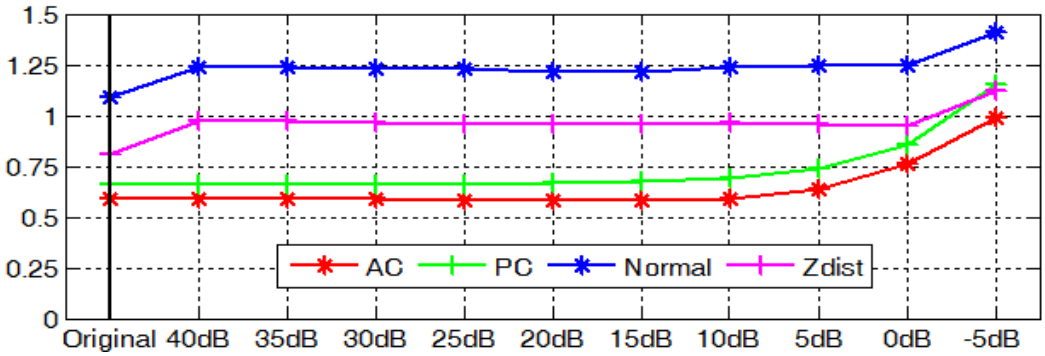

Figure 3.15. Localization errors for the AC, PC, and MSP in millimeters, degrees, or voxels with additive Gaussian noise of zero mean and varying variances.

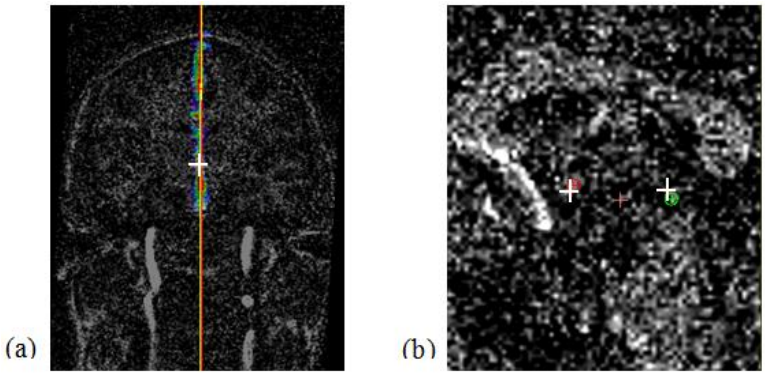

Figure 3.16. Results for a degraded test volume with SNR = -5 dB. Left panel (a) shows the image resampled in the AC-PC space with the response map for the MSP overlaid on top. The yellow line represents the ground truth MSP and red line the detected plane. Right panel (b) shows a zoomed sagittal view of the image, with the white cross being the projection of the ground truth AC/PC, and the red and green point the detected AC and PC respectively.

Robustness with respect to Rotation

To study the sensitivity of our models to rotation, we rotate each test volume with respect to the center of the image around each of the x-, y-, and z-axis separately with angles varying

from 0˚ to 20˚ in 1˚ intervals. For our images, the x-axis corresponds to the anterior-posterior direction, the y-axis the inferior-superior direction, and the z-axis the left-right direction.

Before running the tests, we slightly adjust our algorithm to handle rotations in test images. At each resolution level, we localize target anatomies as previously described, estimate the rotation angles around the x and y axes based on the current estimate of the MSP, and use these angles to reorient the image for the next level. Localization errors for the AC, PC, and MSP are computed in the original space. Rotation correction is needed because of the limited rotational variations around the x and y axes in our dataset. During image acquisition, the scanner restricts the movement of the patient's head from left to right or anterior to posterior. Patients can only freely move by looking up or down, which corresponds to rotation around the z-axis in our case. Rotation around the z-axis, however, does not affect the normal direction of the MSP. This leads to limited angular variations of MSP in our dataset. Models trained with those subjects do not generalize well with images with large rotations around x and y axes and the above modifications in the methodology are necessary.

With the modified method, the localization accuracy with regard to image rotations in the x-, y-, and z- axis is shown in Figure 3.17, 3.18, and 3.19 respectively. Our algorithm for detecting the AC, PC and MSP is robust to rotations in all three directions, with mean errors up to 0.74 mm for AC, 1.07 mm for PC, 1.25˚ and 1.30 voxel for MSP.

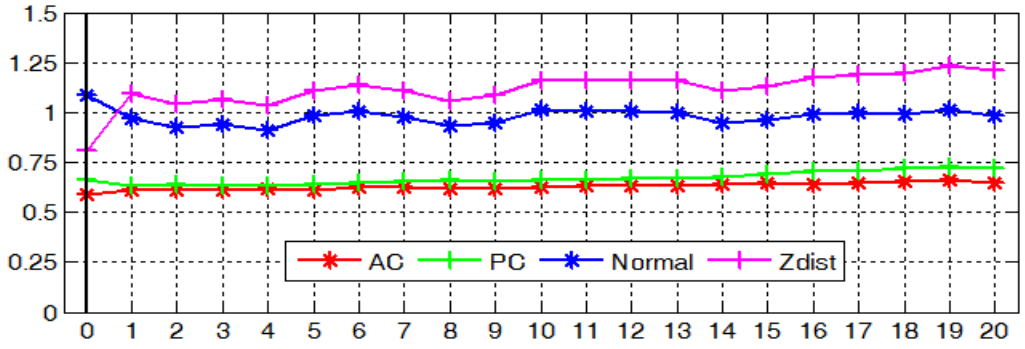

Figure 3.17. Localization errors for the AC, PC, and MSP in millimeters, degrees, or voxels with simulated rotation around the x-axis from 0 to 20 degrees.

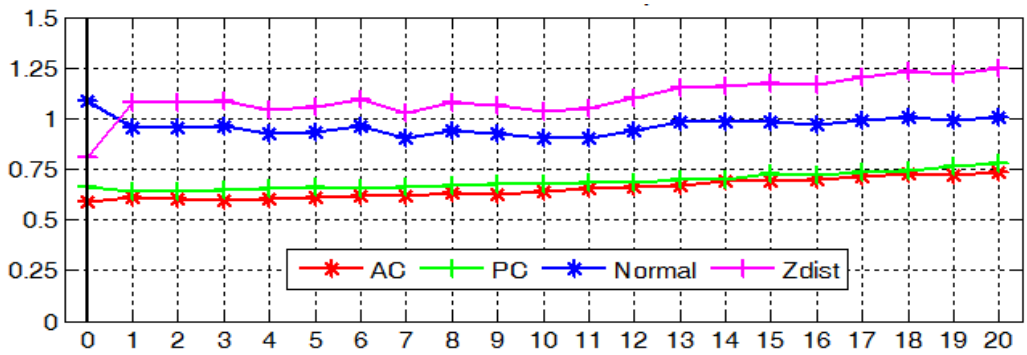

Figure 3.18. Localization errors for the AC, PC, and MSP in millimeters, degrees, or voxels with simulated rotation around the y-axis from 0 to 20 degrees.

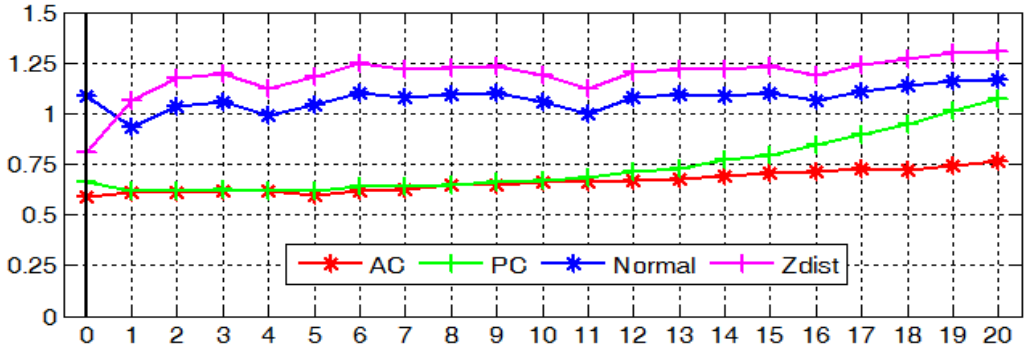

Figure 3.19. Localization errors for the AC, PC, and MSP in millimeters, degrees, or voxels with simulated rotation around the z-axis from 0 to 20 degrees.

3.3.3. Parameter Sensitivity Analysis

We have also investigated the sensitivity of the parameters on the localization accuracy of the AC, PC and MSP. In this study, we select a subset of parameters that may critically affect the results. These parameters include (a) the number of trees in the forest, (b) the size of the node used to terminate the training process, i.e., the number of samples leaf nodes can contain, (c) the number of features to examine per node, and (d) the size of the Gaussian kernel in the mean shift algorithm. Among these parameters, (a)-(c) are used for model creation, i.e., training, and (d) for landmark detection. For each experiment, we test a set of parameter values with the 10 subjects used in section 3.3.2.

To test parameter (a), we plot the out-of-bag (OOB) error rate versus the number of trees for the nine models averaged over the 10 subjects in Figure 3.20. The OOB error is estimated by feeding each tree with the training data left out in the construction of this tree as testing data and calculating the mean square difference between the true probabilities and the predictions. This has been shown to be an unbiased estimate of the test set error [26]. We downscaled the OOB errors for MSP by a factor of 10 for visualization purpose. As more trees are added, the OOB error first rapidly decreases and gradually reaches a plateau with 20, which is the number of trees we have used in section 3.3.1.

For parameter (a), we also assess its direct impact on the localization accuracy by detecting the AC, PC, and MSP with only a subset of trees to make predictions. Localization errors are shown in Figure 3.21. Compared to Figure 3.20, we observe the same trend that introducing more trees help reduce the errors, with a difference that few trees are actually needed. This is because while the OOB plot treats each sample equally, the localization task does not. Samples with low probabilities to be the landmark or in the plane have less effect on the

localization accuracy if their errors are within the tolerance level, as those samples would be disregarded from voting for the final prediction.

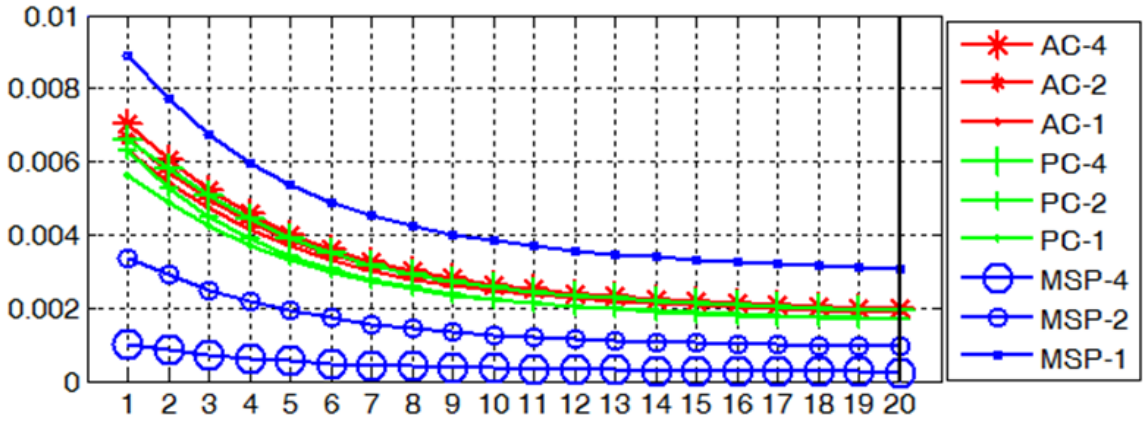

Figure 3.20. OOB errors versus the number of trees for models built for the AC, PC, and MSP at three different resolution levels. Errors for MSP are downscaled by a factor of 10 for display purpose.

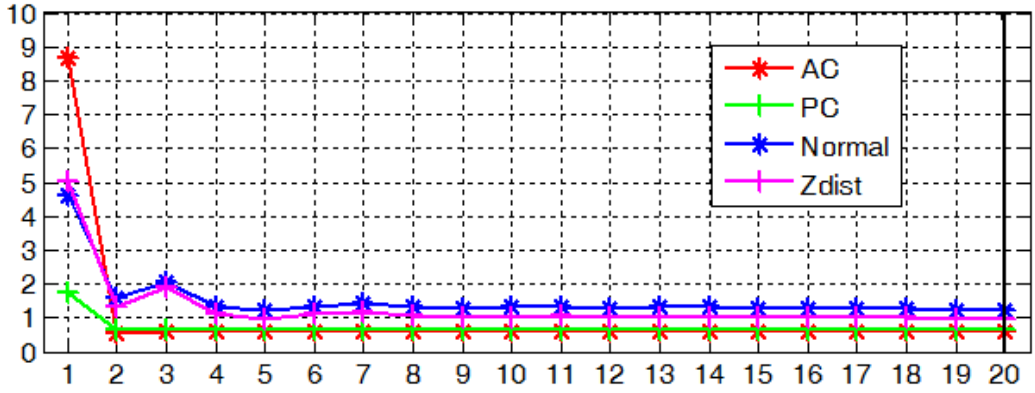

Figure 3.21. Localization errors for the AC, PC, and MSP using models with different number of trees.

To test parameters (b) and (c), for each of the 10 subjects, we re-trained models as described in section 3.3.1 with the same training data but different parameter values. We vary parameter (b) in the interval [1, 20] with a step size of 5. A node size of 1 corresponds to fully grown trees. As the node size increases, the degree of pruning increases. For parameter (c), we test standard values as suggested in [26], i.e., $\sqrt{M}$ and $\log_2 M$ with M being the number of

features for a sample. In our case, these correspond to ~10 and ~50 respectively. We also test a range of values between these and the one we have used in our experiments. We then localize the AC, PC, and MSP using those re-trained models. Sensitivity of the localization errors to parameters (b) and (c) is shown in Figure 3.22 and 3.23 respectively. The values we have used in section 3.3.1 are shown with bold lines. These figures suggest that the size of the node and the number of features examined per node have little impact on the accuracy of our algorithm. While different parameter values may yield slightly better results than those we have used, for example, examining 250 features per node leads to smaller errors for MSP, the differences are marginal, i.e., within 0.1 degree or voxel.

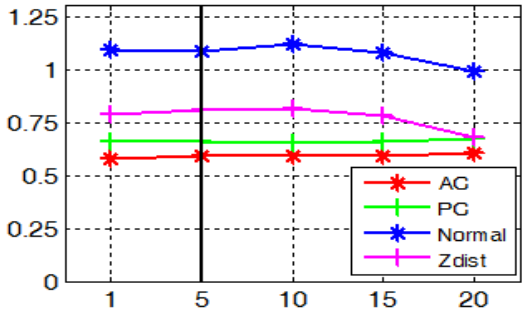

Figure 3.22. Localization errors for the AC, PC, and MSP using models built with different size of the node to terminate training process.

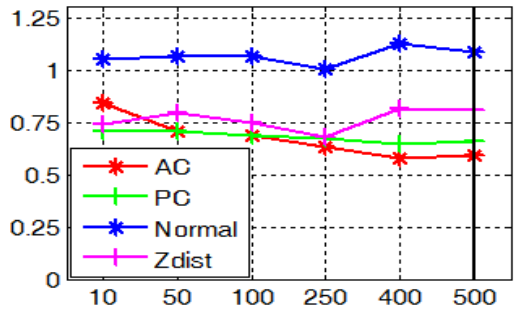

Figure 3.23. Localization errors for the AC, PC, and MSP using models built with different number of features to examine per node.

Figure 3.24 shows the localization errors for the AC and PC when we vary parameter (d). Using a kernel with zero variance corresponds to identifying landmarks as the voxel with the highest probability, as done in [31]. When the variance of the Gaussian kernel increases, errors decline at first and then stabilize when the variance reaches 2. This indicates that the spatial smoothing of the response maps using the mean shift technique improves the AC/PC detection accuracy but the value of the variance has little effect when it is above 2.

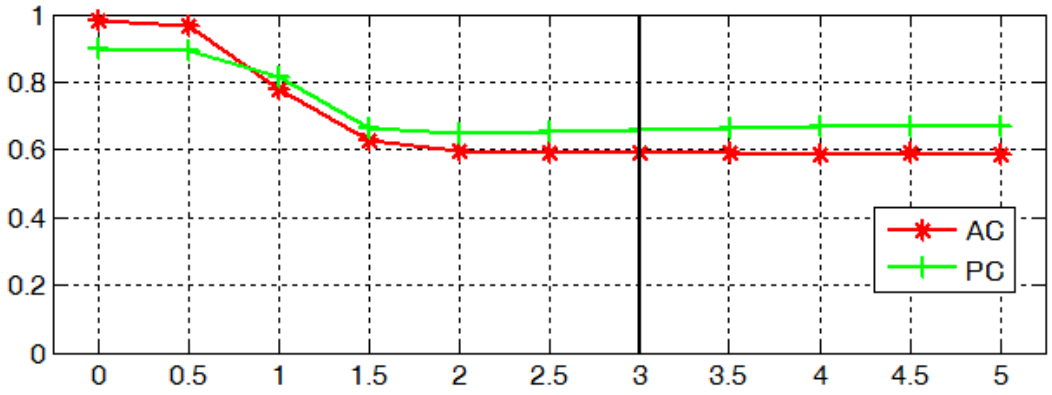

Figure 3.24. Localization errors for the AC and PC in millimeters using Gaussian kernels with zero mean and varying variances.

3.4. Discussions and Conclusions

In this paper, we propose a learning-based method to automatically detect the AC, PC, and MSP in MR T1-w brain scans using random regression forests. We use 20 trees to construct the forest, a number we chose based on the plot of the OOB error versus the number of trees. We have performed a parameter sensitivity analysis, and results indicate that the size of the node used to terminate the training process and the number of features that are examined per node have little impact on the localization accuracy. The features we have used for learning are contextual features generated by randomly displacing vectors. In principle, the number of features could be indefinitely large. To keep the computation practical, only a random set of 2000 features are used.

Recent work by Yaqub *et al.* [46] has suggested that with feature selection the performance of random forests could be enhanced for brain segmentation task in T1-w images, one direction we would like to pursue in our work in the future.

We localize the landmarks and the plane using a semi-local search in the sense that the search is limited to a large region of interest. This is reasonable because all the images in our dataset have similar pose and FOV. If heterogeneous datasets with various orientations, dimensions, and field-of-views are involved, one will need to reorient the image first as we have done when evaluating the sensitivity of our method to rotation to be consistent with the training images or increase the size of the region of interest for training and testing at the coarse level. Our technique is developed on T1-w images. However, extension to other image modalities is straightforward by building models for that particular image set.

We have conducted leave-one-out experiments to validate our method. Results have shown that our approach is accurate and robust, with 0.55 ± 0.30 mm and a maximum of 1.78 mm error for the AC, 0.56 ± 0.28 mm and a maximum of 1.50 mm error for the PC, a $1.08^{\circ} \pm 0.66^{\circ}$ and a maximum of 3.43° angular error in the normal direction, as well as 1.22 ± 0.73 and a maximum of 3.73 voxel of distance error in average z-distance for the MSP.

For the AC and the PC, we have compared our approach to single-atlas-based methods using six well-established nonrigid registration algorithms and also with a model-based approach that has been proposed recently and implemented in the publicly available toolkit YUKI. We have found that our algorithm outperforms four nonrigid registration methods (ABA, DD, F3D, and ART) as well as the toolkit YUKI in terms of accuracy and robustness; the improvements are statistically significant. Other registration methods (LABA and SyN) achieve similar or slightly better accuracy than ours. However, they rely on good affine initialization. In this study we have

manually corrected 9 out of the 100 affine registrations so as not to bias the nonrigid registration results. In an automatic system, such uncorrected affine registration may deteriorate the performance of nonrigid registrations and cause a failure in the AC and PC localization.

For the MSP, we have evaluated our method against a global symmetry-based approach implemented in the toolkit YUKI as well. Our results for both plane normal errors and average z-distances are statistically significantly better than those using YUKI. A thorough comparison to other state-of-the-art methods is nontrivial without any publicly available implementations and is out of the scope of this paper. However, a review of the literature shows that our method compares favorably to existing techniques. Indeed, in [18], Ruppert *et al.* compared their method to Volkau *et al.* [20], Teverovskiy *et al.* [17], and Bergo *et al.* [21] using average z-distance with 20 normal MR T1-w images, and the results were 1.27 ± 0.59, 1.84 ± 1.14, 1.46 ± 0.81, 1.46 ± 0.93 voxel respectively. With a much larger clinical dataset of DBS patients, we achieved results (1.22 ± 0.73) that are highly comparable to the best in [18] (1.27 ± 0.59). In a most recent study which also employed a learning-based schema as discussed earlier [24], Schwing *et al.* reported angular errors in the normal direction of $1.08^{\circ} \pm 0.76^{\circ}$. This is the same mean as ours but their standard deviation is larger.

We have designed a set of experiments to test the robustness of our algorithm against asymmetry of the brain, poor image quality, and variation in rotations across subjects. Results show that our algorithm is tolerant to brain asymmetries such as spherical lesions and tissue deformations without occlusions of the AC/PC. It is also tolerant to high level of imaging noise and, with the modifications described in section 3.3.2, to rotations for the AC, PC, and MSP in any direction.

Another advantage of our approach is the speed. Although registration methods such as

SyN may be marginally more accurate, they generally take at least minutes, if not hours, to run. Our method is fast, taking only about 25 seconds on a standard desktop PC with 4 CPU cores and 8 GB RAM to compute the AC, PC, and MSP all at once. The algorithm is implemented in C++ with parallelization. The computations for model training are done on the Advanced Center for Computing and Research Education (ACCRE) Linux cluster at Vanderbilt University.

We should note that as a supervised learning technique, our method generalizes the information learned from training data to the test data. To achieve the desired accuracy for a test case, patches extracted from this image should look somewhat similar to those in the training set. Fundamentally the performance of our algorithm is limited by the training data we use to build the models. For example, without modifications of our method, the lack of variations in rotation around the anterior-posterior and superior-inferior axes would lead to poor detection of the MSP on test cases with large rotations in those directions. Our algorithm could not manage occlusions of the AC/PC region either, as no such sample appeared in the training data. In order for the model to capture this information, the training set needs to be enriched to handle images with such variations.

There may be some bias towards our method when compared to YUKI, which can be freely downloaded [47]. First, YUKI was not trained on the same images we have used. In addition, it may have used a slightly different definition of the AC, PC, or MSP. For example, the AC could be marked by its center or its posterior edge; the MSP could be defined in other ways than using the falx cerebri. Those factors may account for some of the performance differences between YUKI and our method. To address this issue and permit a thorough comparison between algorithms, it would be beneficial to develop publicly available annotated datasets on which algorithms could be applied.

References

[1] G. Schaltenbrand and W. Wahren, *Guide to the Atlas for Stereotaxy of the Human Brain*. Stuttgart, Germany: Thieme, 1977.

[2] J. Talairach and P. Tournoux, *Co-planar Stereotaxic Atlas of the Human Brain*. Stuttgart, Germany: Thieme, 1988.

[3] P. A. Starr, "Placement of deep brain stimulators into the subthalamic nucleus or globus pallidus internus: technical approach," *Stereotact Funct Neurosurg*, vol. 79, no. 3-4, pp. 118-145, 2003.

[4] W. L. Nowinski and A. Thirunavuukarasuu, "Atlas-assisted localization analysis of functional images," *Med Image Anal*, vol. 5, no. 3, pp. 207-220, 2001.

[5] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," *J Comput Assist Tomo*, vol. 18, no. 2, pp. 192-205, 1994.

[6] Y. Han and H. Park, "Automatic brain MR image registration based on Talairach reference system," in *Proc. ICIP*, 2003, pp. 1097-1100.

[7] S. Joshi, P. Lorenzen, G. Gerig, and E. Bullitt, "Structural and radiometric asymmetry in brain images," *Med Image Anal*, vol. 7, no. 2, pp. 155-170, 2003.

[8] S. Pallavaram, H. Yu, J. Spooner, P.-F. D'Haese, B. Bodenheimer, P. E. Konrad, and B. M. Dawant, "Intersurgeon variability in the selection of anterior and posterior commissures and its potential effects on target localization," *Stereotact Funct Neurosurg*, vol. 86, no. 2, pp. 113-119, 2008.

[9] L. Verard, P. Allain, J. M. Travere, J. C. Baron, and D. Bloyet, "Fully automatic identification of AC and PC landmarks on brain MRI using scene analysis," *IEEE Trans Med Imaging*, vol. 16, no. 5, pp. 610-616, 1997.

[10] K. N. Bhanu Prakash, Q. Hu, A. Aziz, and W. L. Nowinski, "Rapid and automatic localization of the anterior and posterior commissure point landmarks in MR volumetric neuroimages," *Acad Radiol*, vol. 13, no. 1, pp. 36-54, 2006.

[11] G. Zhang, Y. Fu, S. Wang, W. Gao, "Automatic localization of AC and PC landmarks in T2-weighted MR volumetric neuroimages," in *Proc. ICIA*, 2010, pp. 1830-1834.

[12] B. A. Ardekani and A. H. Bachman, "Model-based automatic detection of the anterior and posterior commissures on MRI scans," *Neuroimage*, vol. 46, no. 3, pp. 677-682, 2009.

[13] S. Pallavaram, H. Yu, P.-F. D'Haese, J. Spooner, T. Koyama, B. Bodenheimer, C. Kao, P. E. Konrad, and B. M. Dawant, "Automated selection of anterior and posterior commissures based on a deformable atlas and its evaluation based on manual selections by neurosurgeons," in *Proc. SPIE Med Imaging*, 2007, pp. 65091C-65091C.

[14] P. Anbazhagan, A. Carass, P. L. Bazin, J. L. Prince, "Automatic estimation of midsagittal plane and AC-PC alignment on nonrigid registration," in *Proc. ISBI*, 2006, pp. 828-831.

[15] B. A. Ardekani, J. Kershaw, M. Braun, and I. Kanno, "Automatic detection of the midsagittal plane in 3d brain images," *IEEE Trans Med Imaging*, vol. 16, no. 6, pp. 947–952, 1997.

[16] Y. Liu, R. T. Collins, and W. E. Rothfus, "Robust midsagittal plane extraction from normal and pathological 3D neuroradiology image," *IEEE Trans Med Imaging*, vol. 20, no. 3, pp. 173-192, 2001.

[17] L. Teverovskiy and Y. Liu, "Truly 3D midsagittal plane extraction for robust neuroimage registration," in *Proc. ISBI*, 2006, pp. 860-863.

[18] G. C. S. Ruppert, L. Teverovskiy, C. P. Yu, A. X. Falcao, and Y. Liu, "A new symmetry-based method for midsagittal plane extraction in neuroimages," in *Proc. ISBI*, 2011, pp. 285-288.

[19] M. E. Brummer, "Hough transform detection of the longitudinal fissure in tomographic head images," *IEEE Trans Med Imaging*, vol. 10. No. 1, pp. 74-81, 1991.

[20] I. Volkau, K. B. Prakash, A. Ananthasubramaniam, A. Aziz, and W. L. Nowinski, "Extraction of the midsagittal plane from morphological neuroimages using the kullback-leibler's measure," *Med Image Anal*, vol. 10, no. 6, pp. 863 – 874, 2006.

[21] F. P. G. Bergo, G. C. S. Ruppert, and A. X. Falcao, "Fast and robust mid-sagittal plane location in 3D MR images of the brain," in *Proc. BIOSIGNALS*, 2008, pp. 92-99.

[22] H. Wu, D. Wang, L. Shi, Z. Wen, and Z. Ming, "Midsagittal plane extraction from brain images based on 3D SIFT," *Phys Med Biol*, vol. 59, no. 6, pp. 1367-1387, 2014.

[23] Q. Hu and W. L. Nowinski, "A rapid algorithm for robust and automatic extraction of the midsagittal plane of the human cerebrum from neuroimages based on local symmetry and outlier removal," *Neuroimage*, vol. 20, no. 4, pp. 2153-2165, 2003.

[24] A. Schwing and Y. Zheng, "Reliable extraction of the midsagittal plane in 3d brain mri via hierarchical landmark detection," in *Proc. ISBI*, 2014, pp. 1-4.

[25] A. W. Toga and P. M. Thompson, "Mapping brain asymmetry," *Nat Rev Neurosci*, vol. 4, no. 1, pp. 37–48, 2003.

[26] L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, pp. 5-32, 2001.

[27] M. A. Dabbah, *et al.*, "Detection and location of 127 anatomical landmarks in diverse CT datasets," in *Proc. SPIE Med Imaging*, 2014, pp. 903415-903415.

[28] C. Lindner, S. Thiagarajah, J. M. Wilkinson, T. A. Consortium, G. A. Wallis, and T. F. Cootes, "Fully automatic segmentation of the proximal femur using random forest regression voting," *IEEE Trans Med Imaging*, vol. 32, no. 8, pp. 181-189, 2013.

[29] B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu, "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in *Proc. MICCAI*, 2012, pp. 590-598.

[30] M. Yaqub, A. Kopuri, S. Rueda, P. B. Sullivan, K. McCormick, and J. A. Noble, "A constrained regression forests solution to 3D fetal ultrasound plane localization for longitudinal analysis of brain growth and maturation," in *Proc. MLMI*, 2014, pp. 109-116.

[31] Y. Liu and B. M. Dawant, "Automatic detection of the anterior and posterior commissures on MRI scans using regression forests," in *Proc. EMBC*, 2014, pp. 1505-1508.

[32] P.-F. D'Haese, S. Pallavaram, R. Li, M. S. Remple, C. Kao, J. S. Neimat, P. E. Konrad, and B. M. Dawant, "CranialVault and its CRAVE tools: a clinical computer assistance system for deep brain stimulation therapy," *Med Image Anal*, vol. 16, no. 3, pp. 744-753, 2012.

[33] M. Baehr and M. Frotscher, *Duus' topical diagnosis in neurology: anatomy, physiology, signs, symptoms*, 5th ed., Stuttgart, Germany: Thieme, 2012.

[34] O. Pauly, B. Glocker, A. Criminisi, D. Mateus, A. M. Moller, S. Nekolla, and N. Navab, "Fast multiple organ detection and localization in whole-body MR Dixon sequences," in *Proc. MICCAI*, 2011, pp. 239-247.

[35] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans Pattern Anal Mach Intell*, vol. 17, no. 8, pp. 790-799, 1995.

[36] S. Pallavaram, B. M. Dawant, T. Koyama, H. Yu, J. Neimat, P. E. Konrad, and P.-F. D'Haese, "Validation of a fully automatic method for the routine selection of the anterior and posterior commissures in magnetic resonance images," *Stereotact Funct Neurosurg*, vol. 87, no. 3, pp. 148-154, 2009.

[37] F. Maes, F. A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans Med Imaging*, vol. 16, no. 2, pp. 187-198, 1997.

[38] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int J Comput Vis*, vol. 24, no. 2, pp. 137-154, 1997.

[39] G. K. Rohde, A. Aldroubi, and B. M. Dawant, "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Trans Med Imaging*, vol. 22, no. 11, pp. 1470-1479, 2003.

[40] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: efficient non-parametric image registration," *Neuroimage*, vol. 45, no. 1, pp. 61-72, 2009.

[41] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Med Image Anal*, vol. 12, no. 1, pp. 26-41, 2008.

[42] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Comput Meth Prog Bio*, vol. 98, no. 3, pp. 278-284, 2010.

[43] B. A. Ardekani, S. Guckemus, A. Bachman, M. J. Hoptman, M. Wojtaszek, and J. Nierenberg, "Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans," *J Neurosci Methods*, vol. 142, no. 1, pp. 67-76, 2005.

[44] Y. Liu, P.-F. DH'aese, and B. M. Dawant, "Effects of deformable registration algorithms on the creation of statistical maps for preoperative targeting in deep brain stimulation procedures," in *Proc. SPIE Med Imaging*, 2014, pp. 90362B-90362B.

[45] Z. Han, "Effect of nonrigid registration algorithms on the analysis of brain MR images with deformation based morphometry," Ph.D. dissertation, EECS Dept., Vanderbilt Univ., TN, 2011.

[46] M. Yaqub, M. Javaid, C. Cooper, and A. Noble, "Investigation of the role of feature selection and weighted voting in random forests for 3D volumetric segmentation," *IEEE Trans Med Imaging*, vol. 33, no. 2, pp. 258-271, 2014.

[47] http://www.nitrc.org/frs/?group_id=90

CHAPTER IV

EFFECTS OF DEFORMABLE REGISTRATION ALGORITHMS ON THE CREATION OF STATISTICAL MAPS FOR PREOPERATIVE TARGETING IN DEEP BRAIN STIMULATION PROCEDURES[3]

Yuan Liu, Pierre-François D'Haese, and Benoit M. Dawant

Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37232 USA

[3] [This chapter has been published in *Proceedings of SPIE: Medical Imaging*, San Diego, CA, USA, pp. 90362B-1-90362B-8, 2014.]

Abstract

Deep brain stimulation, which is used to treat various neurological disorders, involves implanting a permanent electrode into precise targets deep in the brain. Accurate preoperative localization of the targets on preoperative MR sequence is challenging as these are typically located in homogenous regions with poor contrast. Population-based statistical atlases can assist with this process. Such atlases are created by acquiring the location of efficacious regions from numerous subjects and projecting them onto a common reference image volume using some normalization method. In previous work, we presented results concluding that nonrigid registration provided the best result for such normalization. However, this process could be biased by the choice of the reference image and/or registration approach. In this paper, we have qualitatively and quantitatively compared the performance of six recognized deformable registration methods at normalizing such data in poor contrasted regions onto three different reference volumes using a unique set of data from 100 patients. We study various metrics designed to measure the centroid, spread, and shape of the normalized data. This study leads to a total of 1800 deformable registrations and results show that statistical atlases constructed using different deformable registration methods share comparable centroids and spreads with marginal differences in their shape. Among the six methods being studied, Diffeomorphic Demons produces the largest spreads and centroids that are the furthest apart from the others in general. Among the three atlases, one atlas consistently outperforms the other two with smaller spreads for each algorithm. However, none of the differences in the spreads were found to be statistically significant, across different algorithms or across different atlases.

4.1. Introduction

In Deep brain stimulation (DBS) procedures, a permanent implant is placed in small deep brain regions of interest such as the subthalamic nucleus (STN), usually referred to as targets. Clinically, this is achieved in a two stage process: an approximate location of the targets is identified preoperatively using current imaging techniques, and then refined during the surgical procedure using responses to stimulation and/or micro-electrode recordings. Preoperative localization of the targets can be done either by localizing the target directly on the patient scan or, as current magnetic resonance (MR) imaging techniques fail to provide sufficient contrast around the targets, using the assistance of atlas-based approaches [1-9]. Printed and digitized anatomical atlases [10-11] as well as population-based statistical atlases [1-4] have been adopted to provide such assistance [1-9]. Creation of statistical atlases of optimum target location typically consists of three steps: 1) identifying the target positions in many subjects, 2) projecting these points onto the reference volumes, and 3) combining them to produce a statistical representation.

When relying on atlas-based approaches to support preoperative targeting, accurate and robust nonrigid registration of 3D image volumes is crucial [12]. Over the last 20 years, a number of nonlinear registration algorithms have been developed but assessing the relative advantages and disadvantages of these algorithms remains challenging. This is because, as opposed to the rigid body registration problem [13], no universal gold-standard exists and algorithms may perform differently for different tasks. Numerous studies have been conducted to evaluate and compare their performances using intensity-based, segmentation-based, deformation field-based, or landmark-based measurements [14-19]. The most comprehensive comparison work we know of is the evaluation of 14 publicly available nonrigid registration

algorithms on four public datasets [16]. Studies studying specifically preoperative targeting in DBS include Chakravarty *et al.* [18] who tested the accuracy of various warping techniques by comparing surface distances and volume overlap for manually labeled subcortical nuclei as well as Castro *et al.* [19] and Duay *et al.* [6] who relied on the manual segmentation of the STN and measured the difference between the centroids of those segmentations obtained by different registration methods. However, these studies require advanced imaging techniques to provide sufficient contrast around targets and substantial amount of manual work to correctly delineate structures of interest, thus making it difficult to extend to a large scale evaluation.

In this paper, we rely on the final intraoperative position of the implant in DBS. Assuming the optimal placement for the lead, its position is a landmark that can be projected through registration to a new patient to predict the target position. However, because of inter-subject variability in this optimal position and sub-optimal intraoperative placement, a single point is not a reliable predictor. Instead, probability maps of target points needs to be computed from a large population of subjects after normalizing the data to the reference volumes. Centroids, spreads and shapes of the maps are important features of the normalized data that may be affected by the choice of the reference volumes or the registration method used for normalization.

The main goal of this paper is to investigate the effect of registration algorithms on statistical atlases in low contrast regions. To achieve this, we use six well-known deformable registration methods to construct statistical maps as described in [1] and both qualitatively and quantitatively evaluate their differences with respect to their centroids, spreads, and shapes. In addition, to avoid any bias in the evaluation of registrations by relying on one single reference space, we use three different atlases, and by analyzing the results obtained for different atlases,

we study the effect of reference volumes on the quality of the statistical maps. We use a large data repository of DBS cases we have gathered over 10 years [20] and computational resources from the Advanced Center for Computing and Research Education (ACCRE) cluster at Vanderbilt University to automate the processing for a large-scale study. The target position is chosen to be the position at which the surgical team choses to place the implant during the procedure, which we will refer to as Final Implant Position (FEP). In the following sections, we first describe the data set we have created over the years. We then explain the method we use to create statistical atlases using six nonrigid registration algorithms. Finally, we present both qualitative and quantitative evaluations for those statistical atlases, and draw several conclusions that are important in DBS.

4.2. Methods

In this section of the paper, we present our methods for evaluating statistical atlases created by different registration methods. First, we discuss the data we use, and then detail our methods for applying different registration methods, for creating the statistical maps, and for designing metrics for comparison.

4.2.1. Data

Over the last decade, we have created a large data repository for DBS surgeries. This repository now contains image volumes for over 1200 subjects (typically preoperative CT and MR volumes and postoperative CT volumes) acquired both at our institution and at collaborating sites. In addition to the images, this repository also contain information such as the Anterior Commissure (AC) and Posterior Commissure (PC) points selected by the surgeons when planning the

procedure, the preoperative target point, the position at which the implant was placed, intraoperative recordings, etc. To conduct the study discussed in this article, we have randomly selected 100 volumes all targeting the STN from this repository. For each of these volumes, we obtain the FEPs identified as the mid-point of the Medtronic 3389 implant. In order to filter patients with extremely unusual anatomy or who were misimplanted, we rely on the FEPs' positions in a coordinate system defined by the AC, the PC, and the midsagittal plane as described in Chapter III, which is frequently used as an indirect targeting approach. The origin of this coordinate system is the mid-point between the AC and the PC, also known as the mid-commissural point. Because of individual variations in the location of the STN relative to this origin, this target coordinate position is not constant across patients. We have only included in our data set patients whose FEPs are within a reasonable distance from the average standardized AC-PC-based target coordinates for STN, which is 12 mm lateral, 3 mm posterior, and 4 mm inferior to the mid-commissural point as reported in [21].

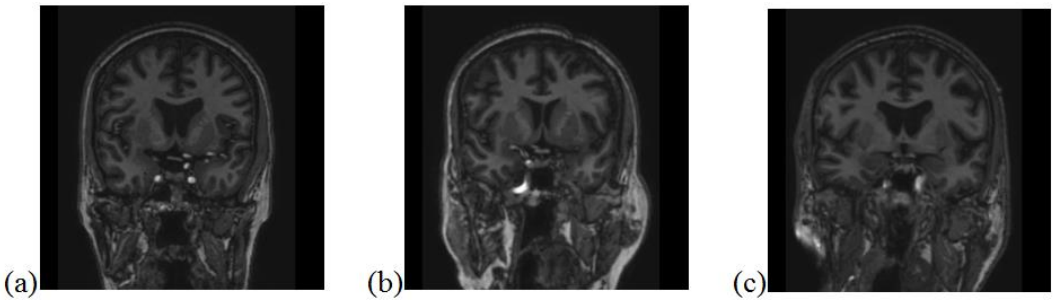

Figure 4.1. The same corresponding coronal slices of the atlases used in this study. From left to right: atlas 1, 2 and 3 respectively.

Among the selected 100 patients, 13 underwent unilateral implantations and the remaining 87 underwent bilateral implantations. This results in 95 FEPs for the left side and 92 for the right side. Three additional reference volumes are chosen as atlases. Among these three

atlases, atlas 2 has a smaller ventricular size than atlas 1 and 3. Corresponding coronal slices in these atlases are show in Figure 4.1. All images in our data set, acquired with IRB approval, are T1-weighted (T1-w) sagittal MR image volumes with approximately 256×256×170 voxels and 1 mm in each direction, acquired with the SENSE parallel imaging technique (T1W/3D/TFE) on a 3 Tesla Phillips scanner (TR = 7.92 ms, TE = 3.65 ms).

4.2.2. Statistical Atlas Creation

For each patient volume, the FEPs are projected from these volumes onto each atlas using the transformation that registers the volume to the atlas. Registrations from patients to atlases are performed in two steps. Each volume is first pre-aligned to each atlas using an affine registration algorithm that uses Mutual Information [22-23] as its similarity measure. As the goal of this study is to compare the nonrigid component of the registration, this step is common to all methods and the results are visually checked and manually corrected to avoid any bias in the study. When a failure is observed, the algorithm is initialized manually and the process repeated until the results are deemed satisfactory. This is followed by the estimation of six dense deformation fields, computed using the six selected registration algorithms that are briefly described below.

Nonrigid Registration Algorithms

The six registration algorithms we have compared in this paper are: 1) the Adaptive Bases Algorithm (ABA) [24], 2) the Adaptive Bases Algorithm with bounding box (LABA) [24], 3) the Diffeomorphic Demons Algorithm (DD) [25], 4) the Fast Free Form Deformation (F3D) [26], 5) Symmetric Normalization (SyN) [27], and 6) the Automatic Registration Tools (ART) [17]. Table 4.1 presents a detailed summary of the differences between these algorithms. More

specifically, LABA is a variation of ABA which limits the computation of the similarity measure within a local area. This local area is defined by a square-shaped bounding box that covers ventricles, basal ganglia structures and the top of the pons. Unlike other algorithms which treat all regions of the brain equally, ABA automatically detects regions of misregistration and places radial basis functions only at those nodes. Without constraining the region of optimization ABA is more likely to focus more on the sulcal area where the gradient of the similarity measure with respect to the coefficients of the basis functions, which is used to localize misregistered regions, is larger. The bounding box is used to limit the region over which the density of basis function is adapted.

Table 4.1. Comparison of deformable registration methods

Method	Deformation Model	Similarity Measure	Regularization	Inverse Deformation Calculation
ABA	Radial basis functions	NMI of whole brain	Transformation symmetry; Jacobian threshold	Yes
LABA	Radial basis functions	NMI in bounding box	Transformation symmetry; Jacobian threshold	Yes
DD	Diffeomorphic optical flow	SSD of whole brain	Gaussian smoothing	No
SyN	Symmetric diffeomorphism	CC of whole brain	Gaussian smoothing; Transformation symmetry	Yes
F3D	Cubic B-splines	NMI of whole brain	Bending energy	No
ART	Homeomorphism	NCC of whole brain	Gaussian smoothing	No

NMI – normalized mutual information, SSD – sum of square differences, CC – cross correlation, NCC – normalized cross correlation

Each algorithm is equipped with a set of parameters that needs to be adjusted for a particular dataset. To do so, we randomly select 10 pairs of volumes to register. Starting from the parameters suggested in a method's user manual or past papers, we test a series of parameter sets on these 10 registrations by adjusting one parameter at a time, i.e., increasing the number of iterations per level, changing the weights for the regularization term, etc. Each time registration quality is assessed visually for global and local alignment (skull, major sulci, and ventricle

aligned correctly) and topological correctness of the deformation field (no self-folding or tearing, indicated by nonnegative jacobian). The parameter set that best fulfills the two objectives is selected for the following experiment.

Inverse Deformation Field Calculation

Warping one image volume onto one atlas requires establishing a transformation that maps one atlas voxel onto a location in the other volume; we refer to this transformation as the forward transformation or forward deformation field. Projecting a point from a volume onto the atlas requires the inverse of this transformation, which we refer to as the inverse transformation or inverse deformation field. While some algorithms compute both others do not. When they do not, the inverse deformation field needs to be estimated from the forward deformation field. Among the algorithms we have evaluated, ABA, LABA, and SyN generate both the forward and the inverse deformation fields. More specifically, ABA and LABA compute both transformations simultaneously and iteratively adjust them to keep bijective transforms of each other at the end of each optimization level. SyN defines a variational energy that explicitly divides the image registration diffeomorphisms into two halves such that the source and target images contribute equally to the path and optimizes from both directions to this virtual midpoint in the space of diffeomorphic transformations. For the other algorithms we use a method available in ITK called itkInvertDisplacementFieldFilter that iteratively computes the inverse field from the forward field by minimizing the inverse consistency error until a certain number of iterations are reached [28].

Point Projection and Aggregation

Once the transformations are obtained, the FEPs are projected from each individual brain volume onto each of the three atlases. This results in 1710 projections (95 points $\times$ 6 methods $\times$ 3

atlases) for FEP on the left side and 1656 projections (92 points × 6 methods × 3 atlases) on the right side that form two clouds of points, one for each side. Those clouds of points, localized in specific reference spaces, are the statistical atlases, often used for guidance for target prediction. For each anatomic atlas, i.e., reference volume, six statistical atlases are created, one for each of the six different registration algorithms. Besides the point clouds, we also build probabilistic maps associated with each statistical atlas. To do this we associate a sphere of 1 mm radius with each point, sum the contribution of all points and normalize this sum to be 1 to represent probability density functions. We then compare the clouds of points as well as the probability maps generated by different registration methods in each of the three reference spaces.

4.2.3. Evaluation

For the evaluation of the statistical atlases, we make two assumptions with regards to the FEPs: 1) during the procedure, the surgical team is able to place the lead at the optimal position; 2) although it is unknown whether or not the placement of the lead should be at the exact same anatomical location for every patient, their positions should be close for patients with the same type of disease. With these assumptions, FEPs should be mapped approximately onto the same location in the atlas. Hence the ideal statistical atlas should be a tight cluster of points that covers some region of the STN.

For qualitative comparison, we visualize the probabilistic maps built for each algorithm as color maps. For each map, we also generate an isosurface using the marching cubes algorithm [29] at a probability of 0.3 in order to compare their shapes. The threshold to extract isosurfaces is chosen empirically to remove outliers with low probabilities while preserving the shape of the

maps. As this small region of interest only contains several voxels, the isosurface is then smoothed for visualization purposes.

To quantitatively measure differences between these statistical atlases, we use two features. The first one is the centroid of the point clouds. Distances between centroids computed by different registration methods indicate differences in their behavior. Second, we compute the absolute deviations for all points in the point cloud and compare the mean and standard deviation of absolute deviations across all point clouds. The absolute deviation is defined as the Euclidean distance between each point and the centroid of the point cloud it belongs to. As we assume that FEPs represent the optimal placement of a lead with respect to a patient's anatomy and that this position should not vary substantially from patient to patient, in a large study as ours we expect that the more robust and accurate the registration algorithm is, the tighter its FEP cluster, and thus the smaller the corresponding mean and standard deviation of absolute deviations.

Besides comparing these features, we perform a one-way analysis of variance (ANOVA) test at a 5% significance level for each atlas at the left and right side to determine whether the means of absolute deviations are the same across all registration algorithms. We also perform the same statistical test for each algorithm on the left and right side to determine whether the means of absolute deviations are the same across all atlases.

4.3. Results

In this section we first show some qualitative results to visually show differences between the statistical atlases. This is followed by a detailed quantitative comparison.

4.3.1. Qualitative Visualization

Figure 4.2 displays the FEP probabilistic maps with their positions relative to the STN in the same slice overlaid on top of the T1-w image for atlas 1. This slice is chosen to contain the midpoint between the AC and the PC and has also been reformatted to be perpendicular to the AC-PC line. At least visually, the maps generated by different registration methods are similar in shapes. One notices, however, that there are more outliers and that the spread is larger for Diffeomorphic Demons. Figure 4.3 shows the 3D isosurfaces created from those maps as well as the contours in the coronal view for all three atlases. As can be observed, most of the contours produced by different registration algorithms share similar shapes, are localized in relatively the same location and cover a similar region of the STN. The blue contours that represent the isosurfaces of the probabilistic maps obtained with Diffeomorphic Demons are shifted compared to the others, most obviously for atlas 2 both on the left side and the right side.

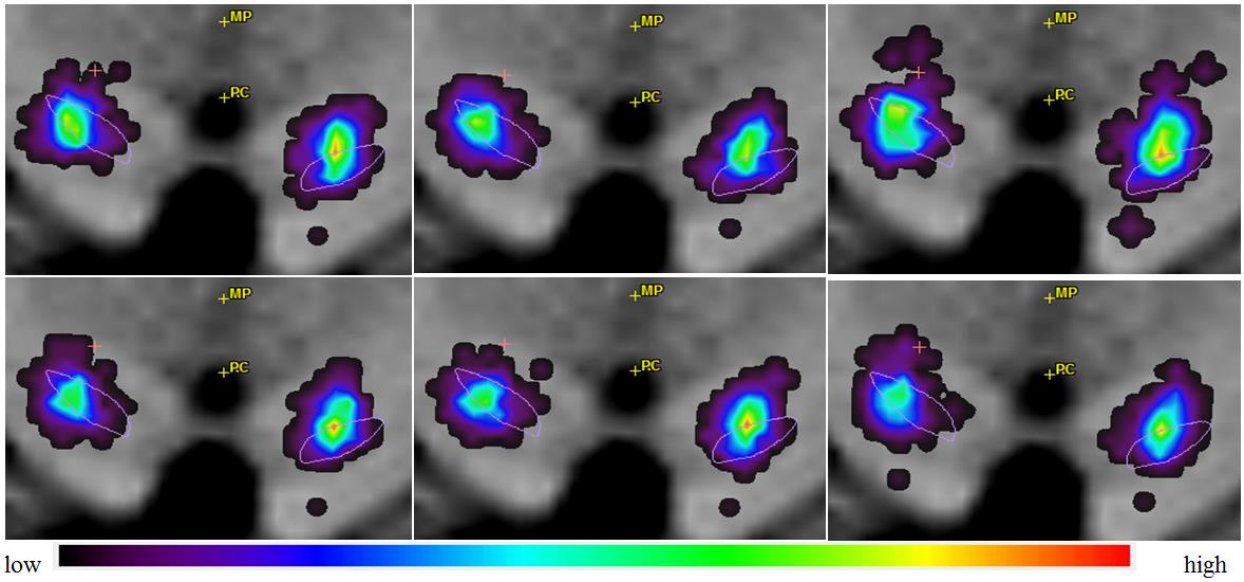

Figure 4.2. FEP probabilistic maps using different registration methods on atlas 1, with STN (purple) drawn on top. From left to right, top row: ABA, LABA, DD; bottom row: SyN, F3D, ART.

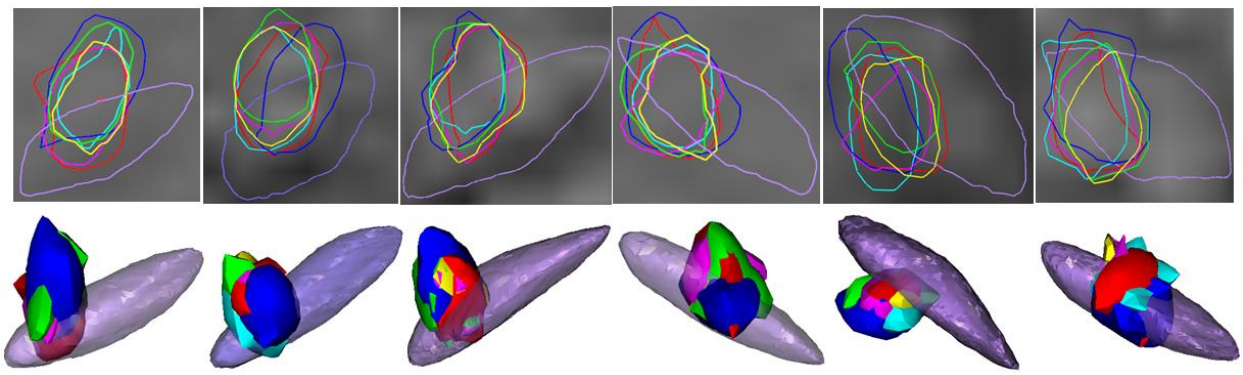

Figure 4.3. Smoothed isosurfaces of the probabilistic maps. Top row is the 2D contours in a coronal slice and bottom row is its 3D visualization. From left to right: left side for atlas 1, 2, and 3, and right side for atlas 1, 2, and 3, respectively. Semi-transparent light purple structure represents the STN. Structures of other colors represent isosurfaces of different methods. Red: ABA; green: LABA; blue: DD; yellow: SyN; magenta: F3D; cyan: ART.

4.3.2. Quantitative Comparison

Table 4.2 reports the Euclidean distances between the centroids of the FEP point clouds obtained with different registration methods for atlas 1, 2, and 3. Centroid differences between methods are all within 1 mm. The centroids obtained by Diffeomorphic Demons are furthest apart from the rest, in average 0.69 mm away. This is consistent with our observations from the qualitative visualizations in Figure 4.2 and 4.3.

Table 4.2. Euclidean distances of centroids between methods in millimeters for the three atlases

	ABA	LABA	DD	SyN	F3D	ART
ABA	0, 0, 0	0.29, 0.36, 0.37	0.33, **0.96**, **0.85**	0.20, 0.06, 0.30	0.33, 0.37, 0.26	0.23, 0.29, 0.43
LABA	0.22, 0.42, 0.35	0, 0, 0	**0.60**, **0.75**, **0.54**	0.09, 0.30, 0.15	0.23, 0.45, 0.16	0.33, 0.31, 0.28
DD	**0.79**, **0.61**, **0.93**	**0.59**, **0.88**, **0.83**	0, 0, 0	**0.52**, **0.94**, **0.61**	**0.56**, **0.80**, **0.64**	0.38, **0.72**, **0.58**
SyN	0.23, 0.38, 0.30	0.05, 0.29, 0.32	**0.60**, **0.94**, **0.65**	0, 0, 0	0.24, 0.37, 0.03	0.29, 0.37, 0.18
F3D	0.16, 0.30, 0.17	0.25, 0.22, 0.29	**0.70**, **0.72**, **0.83**	0.26, 0.27, 0.27	0, 0, 0	0.31, 0.16, 0.21
ART	0.50, 0.44, 0.47	0.39, 0.46, 0.24	0.49, **0.74**, **0.62**	0.42, 0.38, 0.26	0.46, 0.26, 0.41	0, 0, 0

The lower left triangle, shaded with grey, represents calculations for the left side FEPs; the upper right triangle without shading shows the same for the right side. In each cell from left to right the centroid distances are for atlas 1, 2, and 3 respectively. Values larger than 0.5 mm are marked in red bold.

The mean and standard deviation of the absolute deviations of FEPs are shown in Figure 4.4. The means of the absolute deviations range from 2.2 to 2.8 mm for all algorithms in each of the atlases, and their standard deviations is within 1.3-1.65 mm. Differences between those measurements are marginal among the tested registration methods, with Diffeomorphic Demons having the largest mean spreads for atlas 1 and 3. An interesting observation is that mean and standard deviation of the spread for atlas 2 are consistently smaller than those for atlas 1 and 3 across algorithms, with an average decrease of 0.26 mm in mean and 0.14 mm in standard deviation. The results of statistical tests using AVONA further show that differences between mean spread values are not statistically significant from each other for the left and the right side at a 5% significance level, either across different algorithms or across different atlases.

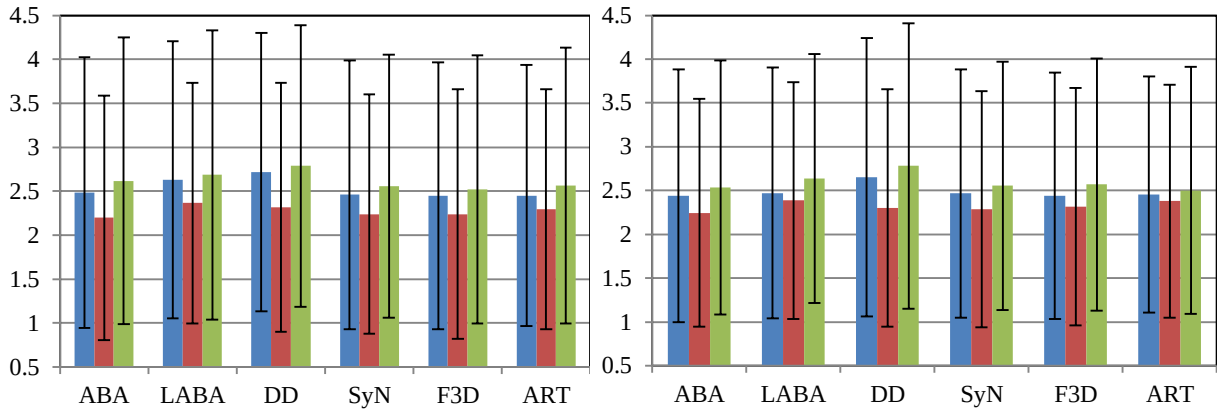

Figure 4.4. Mean and standard deviation of absolute deviations of FEP for the three atlases over all patients. From left to right: FEP left and right side respectively, with blue representing atlas 1, red representing atlas 2, green representing atlas 3.

4.4. Conclusions

In this paper, six well-known deformable registration algorithms are evaluated for the creation of statistical atlases using FEP. This is achieved by registering 100 patients to each of the three atlases using each of the six methods, which results in 1800 deformable registrations. Through

this study, we have found that differences between algorithms' performance in the region of interest are sub-voxel (sub-millimeter). Qualitatively, probabilistic maps generated from the projected clouds of points all cover a similar region of the STN and differences seem negligible. The quantitative analysis confirms this finding. Among these algorithms, the clouds of points generated using the transformations computed by Diffeomorphic Demons are centered about 0.69 mm away from the others in average and present generally larger spreads than those generated by other algorithms. These differences are small when compared to the voxel dimensions. This study leads us to conclude that the choice of registration algorithm does not impact significantly the creation of statistical atlases in the region of the STN when using clinical MR T1-w sequences.

Regarding the choice of the atlases and as discussed above, Figure 4.4 shows that one atlas performs better than the other two with tighter FEP point clouds. Yet, the mean spreads are not significantly different. We have not yet elucidated why this would be the case. The only obvious difference between atlas 2 and the others is the size of the ventricles. It is possible that many volumes in our data set have small ventricles, thus making atlas 2 more similar to these volumes and facilitating registration but we have not verified this hypothesis.

Finally, for each registration algorithm, we select the required parameters using a series of training volumes so that both registration accuracy and topological correctness are optimized. It is possible that results may be affected by parameter values. Exploring this issue would require running our experiments a large number of times with a range of parameter value. While possible with modern computing resources, it remains computationally expensive.

References

[1] P.-F. D'Haese, E. Cetinkaya, P. E. Konrad, C. Kao, and B. M. Dawant, "Computer-aided placement of deep brain stimulators: from planning to intraoperative guidance," *IEEE Trans Med Imaging*, vol. 24, no. 11, pp. 1469-1478, 2005.

[2] S. Pallavaram, *et al.*, "A new method for creating electrophysiological maps for DBS surgery and their application to surgical guidance," in *Proc. MICCAI*, 2008, pp. 670-677.

[3] C. R. Butson, S. E. Cooper, J. M. Henderson, B. Wolgamuth, and C. C. McIntyre, "Probabilistic analysis of activation volumes generated during deep brain stimulation," *NeuroImage*, vol. 54, no. 3, pp. 2096-2104, 2011.

[4] T. Guo, A. G. Parrent, and T. M. Peters, "Surgical targeting accuracy analysis of six methods for subthalamic nucleus deep brain stimulation," *Comput Aided Surg*, vol. 12, no. 6, pp. 325-334, 2007.

[5] S. Bériault, Y. Xiao, L. Bailey, D. L. Collins, A. F. Sadikot, and G. B. Pike, "Towards computer-assisted deep brain stimulation targeting with multiple active contacts," in *Proc. MICCAI*, 2012, pp. 487-494.

[6] V. Duay, X. Bresson, J. S. Castro, C. Pollo, M. B. Cuadra, and J. P. Thiran, "An active contour-based atlas registration model applied to automatic subthalamic nucleus targeting on MRI: method and validation," in *Proc. MICCAI*, 2008, pp. 980-988.

[7] C. Haegelen, P. Coupé, V. Fonov, N. Guizard, P. Jannin, X. Morandi, and D. L. Collins, "Automated segmentation of basal ganglia and deep brain structures in MRI of Parkinson's disease," *Int J Comput Assist Radiol Surg*, vol. 8, no. 1, pp. 99-110, 2013.

[8] M. Keane, S. Deyo, A. Abosch, J. A. Bajwa, and M. D. Johnson, "Improved spatial targeting with directionally segmented deep brain stimulation leads for treating essential tremor," *J Neural Eng*, vol. 9, no. 4, pp. 046005, 2012.

[9] Y. Xiao, *et al.*, "Atlas-Based segmentation of the subthalamic nucleus, red nucleus, and substantia nigra for deep brain stimulation by incorporating multiple MRI contrasts," in *Proc. IPCAI*, 2012, pp. 135-145.

[10] G. Schaltenbrand and W. Wahren, *Guide to the Atlas for Stereotaxy of the Human Brain*. Stuttgart, Germany: Thieme, 1977.

[11] K. A. Ganser, H. Dickhaus, R. Metzner, and C. R. Wirtz, "A deformable digital brain atlas system according to Talairach and Tournoux," *Med Image Anal*, vol. 8, no. 1, pp. 3-22, 2004.

[12] P-F. D'Haese, S. Pallavaram, C. Kao, J. S. Neimat, P. E. Konrad, and B. M. Dawant, "Effect of data normalization on the creation of neuro-probabilistic atlases," *Stereotact Funct Neurosurg*, vol. 91, no. 3, pp. 148-152, 2013.

[13] J. West, *et al.*, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *J Comput Assist Tomogr*, vol. 21, no. 4, pp. 554-566, 1997.

[14] G. E. Christensen, *et al.*, "Introduction to the nonrigid image registration evaluation project (NIREP)," in *Proc. WBIR*, 2006, pp. 128-135.

[15] P. Hellier, X. Geng, J. G. Kuhl, J. Bruss, T. J. Grabowski, I. A. Pirwani, M. W. Vannier, J. S. Allen, and H. Damasio, "Retrospective evaluation of intersubject brain registration," *IEEE Trans Med Imaging*, vol. 22, no. 9, pp. 1120-1130, 2003.

[16] A. Klein, *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *Neuroimage*, vol. 46, no. 3, pp. 786-802, 2009.

[17] B. A. Ardekani, S. Guckemus, A. Bachman, M. J. Hoptman, M. Wojtaszek, and J. Nierenberg, "Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans," *J Neurosci Methods*, vol. 142, no. 1, pp. 67-76, 2005.

[18] M. M. Chakravarty, A. F. Sadikot, J. Germann, P. Hellier, G. Bertrand, and D. L. Collins "Comparison of piece-wise linear, linear, and nonlinear atlas-to-patient warping techniques: analysis of the labeling of subcortical nuclei for functional neurosurgical applications," *Hum Brain Mapp*, vol. 30, no. 11, pp. 3574-3595, 2009.

[19] F. Castro, C. Pollo, R. Meuli, P. Maeder, O. Cuisenaire, M. B. Cuadra, J-G. Villemure, and J-P. Thiran, "A cross validation study of deep brain stimulation targeting: from experts to atlas-based, segmentation-based and automatic registration algorithms," *IEEE Trans Med Imaging*, vol. 25, no. 11, pp. 1440-1450, 2006.

[20] P.-F. D'Haese, S. Pallavaram, R. Li, M. S. Remple, C. Kao, J. S. Neimat, P. E. Konrad, and B. M. Dawant, "CranialVault and its CRAVE tools: a clinical computer assistance system for deep brain stimulation therapy," *Med Image Anal*, vol. 16, no. 3, pp. 744-753, 2012.

[21] P. A. Starr, "Placement of deep brain stimulators into the subthalamic nucleus or globus pallidus internus: technical approach," *Stereotact Funct Neurosurg*, vol. 79, no. 3-4, pp. 118-145, 2003.

[22] F. Maes, F. A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans Med Imaging*, vol. 16, no. 2, pp. 187-198, 1997.

[23] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int J Comput Vis*, vol. 24, no. 2, pp. 137-154, 1997.

[24] G. K. Rohde, A. Aldroubi, and B. M. Dawant, "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Trans Med Imaging*, vol. 22, no. 11, pp. 1470-1479, 2003.

[25] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: efficient non-parametric image registration," *Neuroimage*, vol. 45, no. 1, pp. 61-72, 2009.

[26] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Comput Meth Prog Bio*, vol. 98, no. 3, pp. 278-284, 2010.

[27] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Med Image Anal*, vol. 12, no. 1, pp. 26-41, 2008.

[28] Insight Segmentation and Registration Toolkit (http://www.itk.org/).

[29] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *Proc. SIGGRAPH*, 1987, pp. 163-169.

CHAPTER V

MULTI-MODAL LEARNING-BASED PREOPERATIVE TARGETING IN DEEP BRAIN STIMULATION PROCEDURES[4]

Yuan Liu and Benoit M. Dawant

Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37232 USA

[4] [This chapter has been published in *Proceedings of IEEE International Conference on Biomedical and Health Informatics*, Las Vegas, NV, USA, pp. 17-20, 2016.]

Abstract

Deep brain stimulation, as a primary surgical treatment for various neurological disorders, involves implanting electrodes to stimulate target nuclei within millimeter accuracy. Accurate preoperative target selection is challenging due to the poor contrast in its surrounding region in MR images. In this paper, we present a learning-based method to automatically and rapidly localize the target using multi-modal images. A learning-based technique is applied first to spatially normalize the images in a common coordinate space. Given a point in this space, we extract a heterogeneous set of features that capture spatial and intensity contextual patterns at different scales in each image modality. Regression forests are used to learn a displacement vector of this point to the target. The target is predicted as a weighted aggregation of votes from various test samples, leading to a robust and accurate solution. We conduct five-fold cross validation using 100 subjects and compare our method to three indirect targeting methods, a state-of-the-art statistical atlas-based approach, and four variations of our method with or without spatial features provided with single or multi-modality images. With an overall error of 2.63 ± 1.37 mm, our method significantly outperforms the indirect targeting ones and improves upon any variations by fusing spatial and image information from both modalities. Our technique matches state-of-the-art registration methods but operates on completely different principles. Both techniques can be used in tandem in processing pipelines operating on large databases or in the clinical flow for automated error detection.

5.1. Introduction

Deep brain stimulation (DBS), which sends electrical impulses to specific deep brain nuclei through implanted electrodes, has become a primary surgical treatment for movement disorders

such as Parkinson's disease (PD) [1]. The key to the success of such procedure is the optimal placement of the final implant at millimetric precision level to produce symptomatic relief with minimum or no adverse effects. Traditionally, this is achieved in two stages: a surgical team 1) selects an approximate target location prior to the surgery via visual inspection of anatomic images and 2) adjusts this position based on electrophysiological activities recorded during the surgery. While inaccuracies in preoperative targeting can be corrected in the intraoperative stage, it may prolong the surgery for hours with the patients awake and increase risk such as intra-cranial hemorrhage [2]. Consequently, accurate preoperative planning is highly desirable.

Over the past few decades, numerous efforts have been made to optimize the preoperative target selection. These can be categorized into indirect and direct targeting. Indirect targeting considers the target position as a fixed point in the stereotactic space defined by visible anatomical landmarks, i.e., the anterior commissure (AC), the posterior commissure (PC) and the midsagittal plane (MSP). Various approaches have been proposed to facilitate this task [3-4]. Despite being commonly used, this targeting strategy is limited by the lack of consensus on an ideal anatomic point as the target [5], as well as a failure to account for anatomical variations across patients, i.e., variable width of the third ventricle. On the other hand, direct targeting aims at localizing the target without relying on fixed coordinates. Due to limited contrast in regular T1-weighted (T1-w) magnetic resonance (MR) images, T2-weighted (T2-w) sequences are often acquired for better visualization of the targets. However, T2-w imaging alone does not provide sufficient or consistent contrast, which may lead to a discrepancy between the target positions identified in T2-w images and the ones localized by means of electrophysiological recordings [6]. Alternatively, direct targeting can be assisted with printed and digitized anatomical brain atlases, histological brain atlases, and probabilistic functional atlases alongside nonrigid registration [7-

10]. Once nonlinearly mapped to individual preoperative brain images, such atlases may provide anatomical or/and functional borders of the target nuclei on standard T1-w images.

A recent validation study shows that, using statistical atlases created from a large population, automatic target prediction matches the accuracy of six manual approaches [11]. Routine target prediction is thus achievable but registration failures, especially when nonrigid registration algorithms are used, happens and are difficult to detect automatically. It is thus desirable to develop robust error detection algorithms. One possible approach is to rely on the analysis of the deformation field, but defining quantitative error detection criteria is challenging. Another approach, which is the one proposed herein, is to develop target localization methods that operate on different principles and thus provide another source of information. In this approach, agreement between sources would increase confidence in the predictions while disagreement would indicate a possible error.

Recently, discriminative machine learning techniques have gained popularity for anatomy localization and segmentation. Considering targets as functional landmarks localized anatomically consistently across subjects, we can formulate this problem into a supervised learning framework. One challenge is that as targets are localized in homogeneous regions, image features extracted from the target may not be discernable enough to distinguish its location from adjacent neighbors. Surrounding structures, however, could be used to infer the target location. In fact, this is the underlying principle employed in indirect targeting, which relates the location of the AC, PC, and MSP to a target point. Castro *et al.* have also shown that segmentations of the lateral ventricle, third ventricle, and interpeduncular cistern are useful to improve targeting accuracy [12]. Here, we propose to use regression forests to tackle this problem. Multi-variant regression forests, which combine random forests with Hough transform,

aim to learn multi-dimensional displacement vectors towards an object through a multitude of decision trees [13]. By aggregating predictions made by various test samples, it allows each sample to contribute to an optimal target position with varying degrees of confidence. This is in direct contrast to indirect targeting or direct targeting as done in [12], which require informative structures to be pre-specified. Recently, this regression forests technique has been applied to detect points to drive an active shape model on 2D radiographs [14], to identify the parasagittal plane in ultrasound images [15], and to initialize a nonrigid deformation field in MR images [16]. These successful applications in medical image analysis show promise for target localization.

In this article, we develop a generic multi-modal learning system using T1-w and T2-w images for preoperative targeting. We first apply a learning-based technique to spatially normalize the images in the stereotactic space as used by indirect targeting, and then employ regression forests to learn the displacement vectors towards the target as the latent variable. Targets predicted by this model are spatially constrained via the use of spatial features as is done in indirect targeting, while also accounting for variability of surrounding structures with varying confidence levels by incorporating multi-contrast contextual information. This is independent of the atlas-based registrations, thus permitting quality assurance when used together with atlas-based methods.

5.2. Methods

5.2.1. Data

In this study, we retrospectively examined 100 PD patients with unilateral or bilateral implantations that target the subthalamic nucleus (STN) from a data repository that gathers

patients' data acquired over a decade of DBS surgeries. Every subject was consented to participate in this study. For each subject, this repository contains the clinical active contact locations that are automatically extracted from the Medtronic four-contact 3389 lead in the patient postoperative CT image and projected onto the corresponding preoperative MR T1-w image using a standard intensity-based rigid registration. We use these points as the ground truth target positions for training and evaluation.

The input image data for this study include a T1-w image and a T2-w image of each subject, all acquired as part of the normal delivery of care. T1-w MR image volumes with approximately $256 \times 256 \times 170$ voxel^3 and 1 mm in each direction, were acquired with the SENSE parallel imaging technique (T1W/3D/TFE) on a 3 Tesla Phillips scanner (TR = 7.92 ms, TE = 3.65 ms). T2-w MR image volumes with approximately $512 \times 512 \times 45$ voxel^3 and typical spatial resolution of $0.47 \times 0.47 \times 2$ mm^3, were acquired with the SENSE parallel imaging technique (T2W/3D/TSE) on the same scanner (TR = 3000 ms, TE = 80 ms).

5.2.2. Learning-based Landmark / Plane Identification for Image Pre-alignment

Before training or testing, we first spatially normalize the images using a learning-based technique as described in Chapter III. To do this, a different set of 56 subjects were selected from the data repository, each with a T1-w image and manual annotations of the AC, PC, and MSP. Random forests were used as a regressor to learn a nonlinear mapping between the contextual features of a point and the probability of this point being the AC/PC or in the MSP. After identifying the AC, PC, and MSP, we compute a rigid transformation from the original image space to the AC-PC space as defined in Chapter III. This AC-PC coordinate system is the stereotactic space used to perform indirect targeting [7]. We rigidly align the T2-w images with

their corresponding T1-w images of the same subject using intensity-based techniques. The normalization transformation is applied to both co-registered sequences and these images are resampled at 1 mm isotropic resolution for subsequent training and testing.

5.2.3. Problem Formulation

We use a voxelwise solution based on regression forests to learn the displacement of a voxel to the target and train one forest for each side of the brain. As there is substantial shape variability in the cerebral cortex across subjects, it may be difficult for the model to relate a point in this area with diverse appearances to a target. To simplify model training, we define a bounding box that roughly covers the deep brain and constrain the samples to be uniformly drawn from this region. This region of interest is illustrated by the yellow box in Figure 5.1.

Each training sample is associated with a heterogeneous high dimensional feature vector and a 3D displacement vector $\vec{d} = [\mathrm{dx}, \mathrm{dy}, \mathrm{dz}]$ from its position to the target. Thousands of features are used, which consist of:

- Spatial features, which are the spatial coordinates (x, y, z) of the voxel in the AC-PC space.
- Multi-modal intensity contextual features, which are the mean intensity differences between two cuboids as used in Chapter III [17]. Those cuboids, as described earlier, are randomly displaced by varying amounts to capture multi-scale textural context variations. We compute a number of such features from each image sequence independently.

Spatial features cluster training samples according to their distances to the target. Models built using these features alone have a similar effect on target prediction compared to the indirect targeting approaches, which can be obtained as a byproduct of the pre-alignment step. Intensity features allow models to capture anatomical variations and adjust target positions accordingly.

Moreover, by extracting such features from each sequence, we account for various appearances of different modalities that may complement each other, as shown in Figure 5.1.

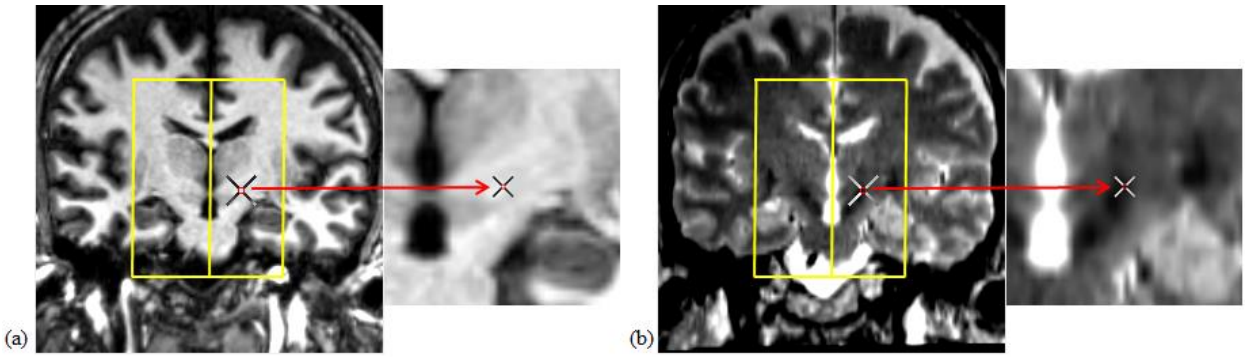

Figure 5.1. Co-registered T1-w and T2-w images of one subject in panel (a) and (b) respectively. The yellow box represents the sampling region, and the cross represents the ground truth target point. The zoomed-in view on the right side of each panel illustrates the different contrast each image sequence reveals around the target.

5.2.4. Regression Forests Training

We use a total of 100 regression trees to construct the forest. To build each tree, a bootstrap of two thirds of training samples is randomly chosen and fed to the root node of the tree. Given the training samples $\{\overrightarrow{f_n}, \overrightarrow{d_n}\}_{n=1}^{N'}$ at a particular node, a feature f^m and a threshold t are selected to split the data, which minimize the mean square error MSE:

$$t, m = \arg\min_{t,m} \left(MSE\left(\{\overrightarrow{d_n}: f_n^m < t\}\right) + MSE\left(\{\overrightarrow{d_n}: f_n^m \geq t\}\right)\right) \quad (5.1)$$

Node splitting is done recursively, and stops when a certain tree depth is reached or the best split threshold cannot be found. At each leaf node, we obtain a multivariate histogram of $\vec{d}$ for the voxels reaching this node. Assuming uncorrelated Gaussian distribution, we store the mean and variance for each dimension as a compressed representation of the histogram.

5.2.5. Regression Forests Testing

During testing, test samples traverse all trees in the forests. Starting from the root node, test samples recursively visit the left or right child based on the binary feature test stored at this split node until a leaf node is reached. To aggregate votes of each test sample from each tree, a range of styles of voting can be used, e.g., casting a single vote per tree as the mean displacement vector, or using multiple votes from the training samples. Cootes *et al.* compared different voting strategies and showed that a single vote per tree performed the best in terms of accuracy and speed [18]. In this paper, we gather a single vote from each tree as the mean displacement vector and measure its confidence based on the variance $[\mathrm{Var}(\mathrm{dx}), \mathrm{Var}(\mathrm{dy}), \mathrm{Var}(\mathrm{dz})]$. Many possible weighting functions can be used to reflect the boosted confidence in the prediction if the variances are small. Here we define the weight of each vote as:

$$\mathrm{w} = e^{-\alpha \cdot (\mathrm{Var}(\mathrm{dx}) + \mathrm{Var}(\mathrm{dy}) + \mathrm{Var}(\mathrm{dz}))} \tag{5.2}$$

To increase robustness, the final prediction is made using the weighted mean shift method with a Gaussian kernel [19]. This point is then projected back to the original image space using the transformation matrix computed before.

5.2.6. Comparison to Other Methods

Provided with the pre-aligned images in the AC-PC space as done in section 5.2.2, indirect targeting approaches can be readily applied. We hereby localize the targets using two widely used indirect targeting methods from the literature. To account for positional variations between our study and the ones published in the literature, we also conduct an experiment that uses training volumes to obtain an indirect estimation. In addition, we compare ours to a state-of-the-

art statistical atlas-based method as validated in [11], which was shown to outperform six manual targeting approaches. The following summarizes the experiments we have performed:

- Indirect targeting using the stereotactic coordinates (±12 mm lateral, -3 mm anterior, -4 mm superior) as the center of the motor territory of the STN [20] (referred to as IND1).
- Indirect targeting using the stereotactic coordinates (±11.8 mm lateral, -2.4 mm anterior, -3.9 mm superior) as a representation of the centroid of active contacts [20] (referred to as IND2).
- Indirect targeting using the stereotactic coordinates of the average active contact position in the training volumes of our dataset (referred to as IND3).
- Statistical atlas-based approach as described in Chapter IV (referred to as SA). Specifically, a statistical atlas is created by performing a series of affine, local affine, and local nonrigid registrations between each training volume and a pre-specified anatomical atlas, and projecting the target locations from the training subject to this reference space. The centroid of those projections is projected onto a test subject through a series of affine and nonrigid transformations and its projection is taken as the predicted target.

To evaluate the importance of each feature type, we compare results obtained with the full feature set and with a subset of features. Calling the method described so far RF-S+T1+T2 for regression forests with spatial and intensity information from T1-w and T2-w images, the other variations we have tested are:

- Regression forests-based targeting using only spatial features (referred to as RF-S). This is similar to IND3 as we only use spatial information derived from training volumes for target prediction. A comparison between this method and RF-S+T1+T2 indicates the importance of image-based features.

- Regression forests-based targeting using spatial features and intensity contextual features extracted from only T1-w or T2-w images (referred to as RF-S+T1 and RF-S+T2 respectively). We keep the dimensionality of the feature vector the same as the one used in RF-S+T1+T2 but only compute mean intensity differences for all contextual features from a single modality. Comparisons between these two methods and RF-S+T1+T2 illustrate the effect of fusing multi-modalities.
- Regression forests-based targeting using only intensity contextual features extracted from T1-w and T2-w sequences as used before (referred to as RF-T1+T2). A comparison between this method and RF-S+T1+T2 shows the effect of spatial features.

5.3. Results

In this study, we conduct a five-fold cross validation, i.e., we use 80 subjects for training and the remaining 20 for testing and we repeat the process five times.

For visualization purpose, we compute response maps as the weighted aggregation of all votes, with predictions at each voxel scaled to the interval [0, 1]. These maps, computed with the various feature sets used in our study, are shown in Figure 5.2 for one test subject. The response maps for RF-S+T1, RF-S+T2, and RF-S+T1+T2 are similar to each other; the peak regions, which indicate the high probability of predictions gathered from all test samples, are tight in those maps. The peak region is even tighter for RF-S, which are almost point estimates. RF-T1+T2, on the contrary, results in widespread predictions with many outliers, due to the lack of spatial regularization as used in the other methods.

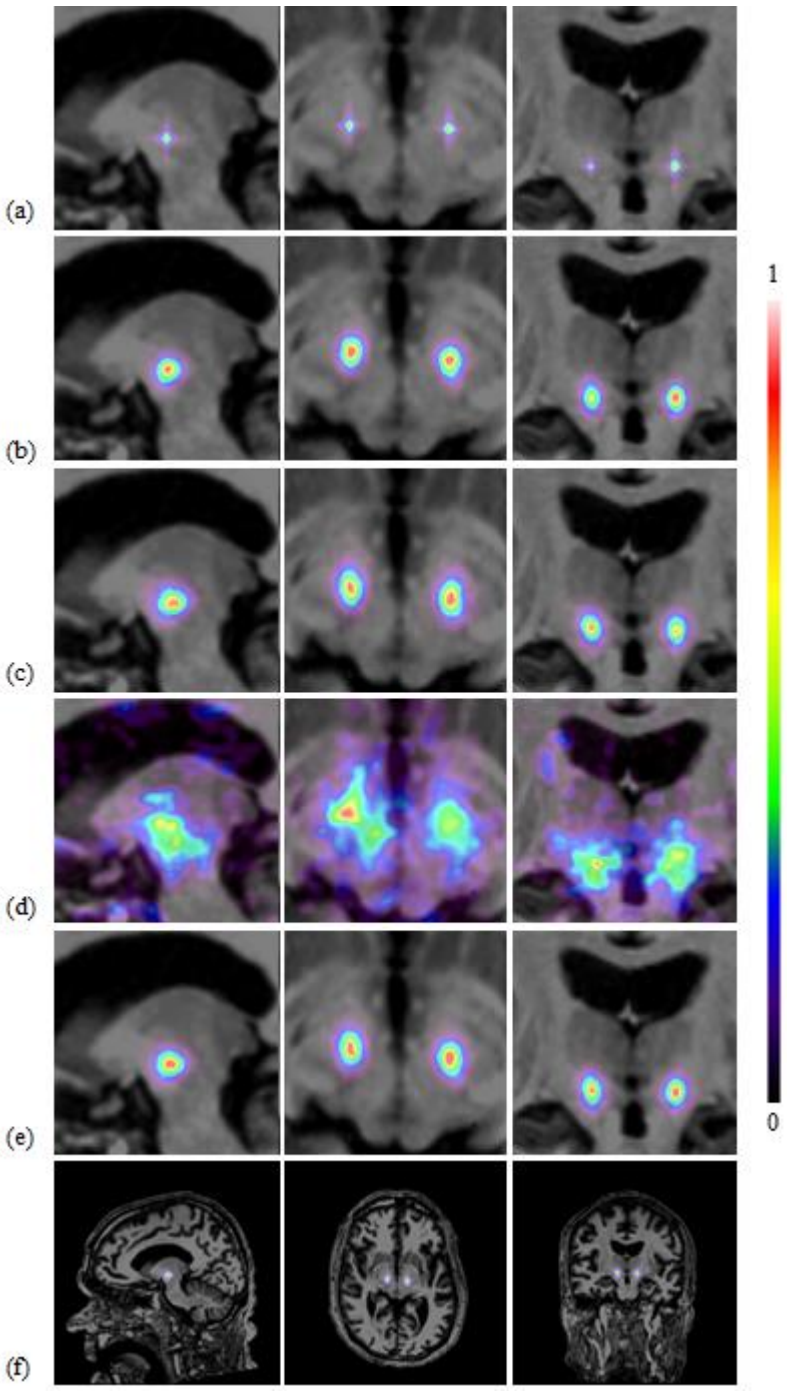

Figure 5.2. Response maps for a test case, using RF-S, RF-S+T1, RF-S+T2, RF-T1+T2, and RF-S+T1+T2 in panels (a)-(e) respectively. Panel (f) is a whole head view of panel (e).

To quantify the targeting accuracy, we define the targeting error as the 3D Euclidean distance between the prediction and the ground truth. A box plot of those errors for all cases is

shown in Figure 5.3, with their mean and standard deviation reported in Table 5.1. Our method performs better than all indirect targeting methods and is comparable to the statistical atlas-based method. Moreover, it results in a reduced overall error compared to: 1) RF-S using only spatial features, which illustrates the importance of image based information; 2) RF-S+T1 and RF-S+T2 using a single modality image, which demonstrates the benefits of fusing multi-contrast information; 3) RF-T1+T2 using only image-based features, which shows the significance of spatial regularization in targeting. In addition, RF-S performs similarly to IND3. This is expected as IND3 can be considered as learning a predictor from training volumes with only spatial information, similar to what RF-S does.

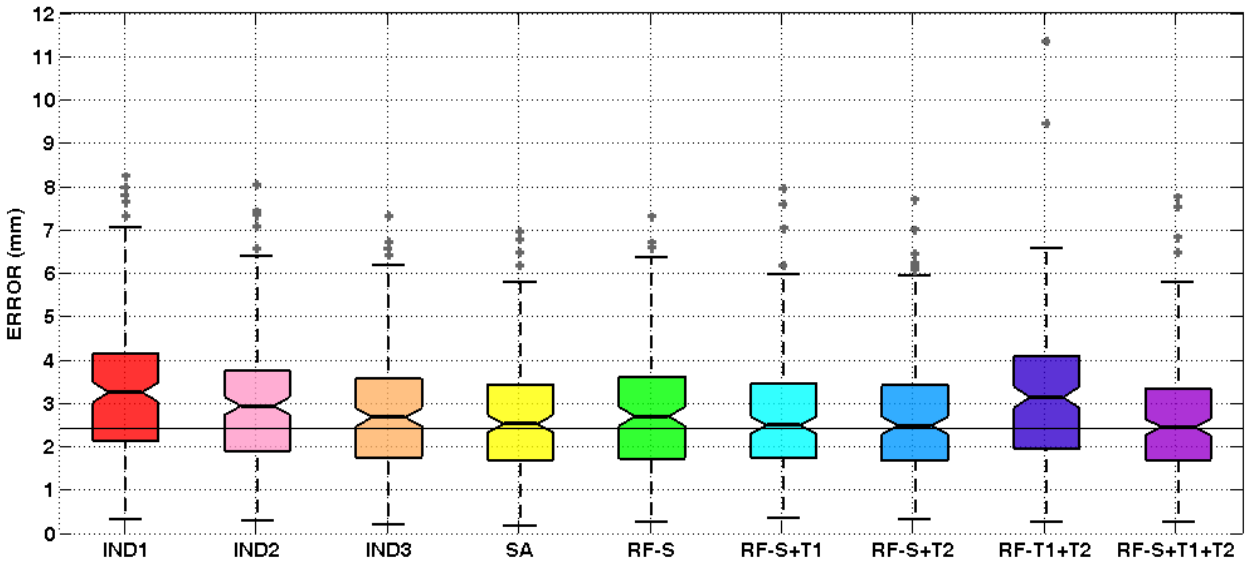

Figure 5.3. Boxplot of targeting errors using different techniques.

Table 5.1. Mean and standard deviations (Std.) of targeting errors using different techniques

Error (mm)	IND1	IND2	IND3	SA	RF-S	RF-S+T1	RF-S+T2	RF-T1+T2	RF-S+T1+T2
Mean	3.26	3.02	2.81	2.68	2.82	2.72	2.68	3.15	2.63
Std.	1.58	1.53	1.36	1.35	1.36	1.37	1.41	1.58	1.37

In addition, we performed one-sided paired Wilcoxon signed-rank tests to test whether the medians of errors using our method are smaller than those using the other methods. The *p*-values are shown in Table 5.2, with values lower than the significance level of 0.05 marked in red bold. As shown in the table, our method significantly reduces the targeting errors compared to all indirect targeting methods, while the difference between our method and the statistical atlas-based method is not found to be significant.

Table 5.2. *P*-values of Wilcoxon tests between the targeting errors using our method and those using other methods

P-values	IND1	IND2	IND3	SA	RF-S	RF-S+T1	RF-S+T2	RF-T1+T2
RF-S+T1+T2	**0.00**	**0.00**	**0.00**	0.25	**0.00**	**0.00**	0.13	**0.00**

5.4. Conclusions

In this paper, we propose a multi-modal learning-based method using regression forests to automatically localize the target in preoperative MR brain scans. By taking advantage of a large dataset of past patients, our approach improves upon indirect targeting by tuning this estimate to patient's individual anatomy and combines the strength of different direct targeting methods by exploiting the multi-contrast information. This is further substantiated by our results with the smallest targeting errors compared to those using spatial features or image features alone, or provided with either modality alone. Our technique also does not require the segmentation of anatomic structures or nonrigid registration. It can be used in tandem with nonrigid registration methods in clinical processing pipelines to develop robust error prediction schemes.

Results from a five-fold cross validation study have shown that our approach is accurate and robust with a 2.63 ± 1.37 mm targeting error. It matches the accuracy of the state-of-the-art statistical atlas-based methods, which outperforms six manual methods as shown in [11]. We

have also found a high correlation ($\rho = 0.83$) between targeting errors made by our approach and the ones made by statistical atlas-based method. This is likely because of the discrepancy between imaging and neurophysiology, which causes both imaging-based methods to induce relatively larger errors for some cases. In the future, it will be useful to explore electrophysiological data to account for this discrepancy. For example, a machine learning model could be built with both imaging data and the electrophysiological recordings of past patients. For the current patient, this model predicts the most plausible target, initially using only image information and later filled in with recordings collected sequentially in the operating room.

Our method is also fast. Once the models are trained on the Advanced Center for Computing and Research Education (ACCRE) Linux cluster at Vanderbilt University, the testing pipeline takes approximately 40 seconds on a standard PC with 4 CPU cores and 8GB RAM. This compares favorably to atlas-based registration methods, which may take from several minutes to hours to run.

References

[1] M. L. Kringelbach, N. Jenkinson, S. L. Owen, and T. Z. Aziz, "Translational principles of deep brain stimulation," *Nat Rev Neurosci*, vol. 8, no. 8, pp. 623-635, 2007.

[2] X. Hu, *et al.*, "Risks of intracranial hemorrhage in patients with Parkinson's disease receiving deep brain stimulation and ablation," *Parkinsonism Re Disord*, vol. 16, no. 2, pp. 96-100, 2010.

[3] Y. Han and H. Park, "Automatic brain MR image registration based on Talairach reference system," in *Proc. ICIP*, 2003, pp. 1097-1100.

[4] Y. Liu and B. M. Dawant, "Automatic detection of the anterior commissure, posterior commissure, and midsagittal plane on MRI scans using regression forests," *IEEE J. Biomed Health Inform*, vol. 19, no. 4, pp. 1362-1374, 2015.

[5] T. Yokoyama, K. Sugiyama, S. Nishizawa, N. Yokota, S. Ohta, S. Akamine, and H. Namba, "The optimal stimulation site for chronic stimulation of the subthalamic nucleus in Parkinson's disease," *Stereotact Funct Neurosurg*, vol. 77, no. 1-4, pp. 61-67, 2001.

[6] E. Cuny, D. Guehl, P. Burbaud, C. Gross, V. Dousset, and A. Rougier, "Lack of agreement between direct magnetic resonance imaging and statistical determination of a subthalamic target: the role of electrophysiological guidance," *J Neurosurg*, vol. 97, no. 3, pp. 591-597, 2002.

[7] G. Schaltenbrand and W. Wahren, *Guide to the Atlas for Stereotaxy of the Human Brain*. Stuttgart, Germany: Thieme, 1977.

[8] K. A. Ganser, H. Dickhaus, R. Metzner, and C. R. Wirtz, "A deformable digital brain atlas system according to Talairach and Tournoux," *Med Image Anal*, vol. 8, no. 1, pp. 3-22, 2004.

[9] J. Yelnik, *et al.*, "A three-dimensional, histological and deformable atlas of the human basal ganglia. I. Atlas construction based on immunohistochemical and MRI data," *Neuroimage*, vol. 34, no. 2, pp. 618–638, 2007.

[10] T. Guo, A.G. Parrent, and T. M. Peters, "Surgical targeting accuracy analysis of six methods for subthalamic nucleus deep brain stimulation," *Comput Aided Surg*, vol. 12, no. 6, pp. 325-334, 2007.

[11] S. Pallavaram, P.-F. D'Haese, W. Lake, P. E. Konrad, B. M. Dawant, and J. S. Neimat, "Fully automated targeting using nonrigid image registration matches accuracy and exceeds precision of best manual approaches to subthalamic deep brain stimulation targeting in Parkinson disease," *Neurosurgery*, vol. 76, no. 6, pp. 756-765, 2015.

[12] F. Castro, C. Pollo, R. Meuli, P. Maeder, O. Cuisenaire, M. B. Cuadra, J-G. Villemure, and J-P. Thiran, "A cross validation study of deep brain stimulation targeting: from experts to atlas-based, segmentation-based and automatic registration algorithms," *IEEE Trans Med Imaging*, vol. 25, no. 11, pp. 1440-1450, 2006.

[13] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Med Image Anal*, vol. 17, no. 8, pp. 1293-1303, 2013.

[14] C. Lindner, S. Thiagarajah, J. Wilkinson, The Consortium, G. Wallis, and T. F. Cootes, "Fully automatic segmentation of the proximal femur using random forest regression voting," *IEEE Trans Med Imaging*, vol. 32, no. 8, pp. 181-189, 2013.

[15] M. Yaqub, M. Javaid, C. Cooper, and A. Noble, "Plane localization in 3D fetal neurosonography for longitudinal analysis of the developing brain," *IEEE J Biomed Health Inform*, vol. 20, no. 4, pp. 1120-1128, 2016.

[16] D. Han, Y. Gao, G. Wu, P. Yap, and D. Shen, "Robust anatomical landmark detection with application to MR brain image registration," *Comput Med Imaging Graph*, vol. 46, pp. 277-290, 2015.

[17] O. Pauly, B. Glocker, A. Criminisi, D. Mateus, A. M. Moller, S. Nekolla, and N. Navab, "Fast multiple organ detection and localization in whole-body MR Dixon sequences," in *Proc. MICCAI*, 2011, pp. 239-247.

[18] T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer, "Robust and accurate shape model fitting using random forest regression voting," in *Proc. ECCV*, 2012, pp. 278-291.

[19] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans Pattern Anal Mach Intell*, vol. 24, no. 5, pp. 603-619, 2002.

[20] P. A. Starr, "Placement of deep brain stimulators into the subthalamic nucleus or globus pallidus internus: technical approach," *Stereotact Funct Neurosurg*, vol. 79, no. 3-4, pp. 118-145, 2003.

CHAPTER VI

GENERATION OF HUMAN THALAMUS ATLASES FROM 7T DATA AND APPLICATION TO INTRATHALAMIC NUCLEI SEGMENTATION IN CLINICAL 3T T1-WEIGHTED IMAGES[5]

Yuan Liu[1], Pierre-François D'Haese[1], Allen T. Newton[2], and Benoit M. Dawant[1]

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37232 USA

[2]Institute of Image Science Department, Vanderbilt University, Nashville, TN 37232 USA

[5] [An earlier version of this chapter, "Thalamic nuclei segmentation in clinical 3T T1-weighted images using high-resolution 7T shape models", has been published in *Proceedings of SPIE: Medical Imaging*, Orlando, FL, USA, pp. 94150E-1-9, 2015.]

Abstract

Accurate and reliable identification of thalamic nuclei is important for surgical interventions and neuroanatomical studies. This is a challenging task due to their small size and low intrathalamic contrast in standard T1-weighted or T2-weighted images. Previously proposed techniques rely on diffusion imaging or functional imaging. These require additional scanning and suffer from the low resolution and signal-to-noise ratio in these images. In this chapter, we present a protocol to build histological-like atlases from high-field MR images and use such atlases to guide the segmentation of thalamic nuclei in standard 3T T1-weighted images. To generate the atlas, we design a set of 7T MR sequences that all together increase intrathalamic contrast to allow manual delineation of the internal substructures. Following this imaging protocol, we manually segment 23 thalamic nuclei on a set of 9 high-resolution brain atlases based on the Morel atlas naming convention. With the aid of these thalamus atlases, we have proposed single-atlas-based, multi-atlas-based, and statistical shape-based approaches to localize the internal nuclei from a given thalamus segmentation. While single-atlas and multi-atlas approaches rely on registration to establish an accurate mapping between the atlas and the subject within the thalamus region, the shape-based approach builds joint shape models of the thalamus and the internal structures, and predicts the target structures by fitting the thalamus to those models. Results of a leave-one-out validation study on nine subjects show that the use of multiple atlases statistically improves upon the single-atlas approach for most nuclei. Segmentations obtained with statistical shape models yield the highest accuracy, with dice coefficients ranging from 0.53 to 0.90 and mean surface errors from 0.34 mm to 0.88 mm for all nuclei averaged across test cases. This suggests the feasibility of using such approach for localizing thalamic substructures. It may have a direct

impact on surgeries such as Deep Brain Stimulation procedures that require the implantation of stimulating electrodes in specific thalamic nuclei.

6.1. Introduction

The thalamus serves as the central relay station for the brain that processes and relays sensory and motor signals between different subcortical regions and the cerebral cortex. It can be divided into several neuronal clusters referred to as nuclei, each possibly subdivided into sub-nuclei. A number of diseases, including Parkinson's disease, multiple sclerosis, and chronic pain syndrome, are associated with some of these nuclei [1-3]. Stimulating by means of Deep Brain Stimulation (DBS) electrodes or ablating by means of radiosurgery specific nuclei, e.g., the ventral intermediate nucleus for movement disorders, can reduce symptoms for these diseases [4]. Identification of intrathalamic structures in individual patients opens new doors for direct anatomical targeting for DBS and investigation of regional thalamic changes related to disease progression. Unfortunately, this cannot be readily achieved by standard T1-weighted (T1-w) or T2-weighted (T2-w) magnetic resonance (MR) sequences due to the low intrathalamic contrast in those images and the small size of the substructures.

To accurately study neuroanatomy, *ex vivo* imaging followed by histological staining remains the gold standard to date. Thalamic nuclei possess distinct cyto- and myelo-architecture that can be differentiated in histological data. Morel *et al.* presented one of the most used histological atlases of the thalamus based on multi-architectonic parcellation in sections parallel or perpendicular to the standard inter-commissural reference plane [5]. In this work, three chemical markers were used to increase contrast across thalamic substructures with different chemical properties. Such atlas consists of a stack of 2D slices overlaid with manual contours for

each specimen. Three-dimensional digitization of those atlases is challenging because of the highly anisotropic resolution and biases towards the individual anatomy [6-7]. To overcome these limitations, Krauth *et al.* proposed to construct a mean model of thalamic structures by averaging over multiple postmortem brains acquired in different orientations [8]. Quality of the model to represent the variety of human anatomy is directly correlated with the number of specimens involved. However, obtaining additional histological specimen is a highly demanding task in terms of resources and expertise. Furthermore, to use such atlases for segmentation, spatial normalization is performed to warp them into individual patients. Most normalization techniques rely on the existence of a good MR reference image, which is often missing in those histological atlases.

Alternatively, diffusion tensor imaging (DTI) and functional MR imaging (fMRI) have been introduced to distinguish the thalamic substructures *in vivo*. The orientation of neuronal pathways, detected by DTI, is relatively consistent within a nucleus. Based on this property, a number of techniques cluster each voxel inside the thalamus with distance metrics designed to measure local tensor inhomogeneity using k-means, spectral clustering, and mean shift algorithms [9-11]. Parcellation of the thalamus can also be achieved by probabilistic tractography according to structural connectivity to different cortical regions [12]. Meanwhile, functional brain organization has been examined using task-related fMRI [13] or resting state fMRI [14], and functional boundaries in the thalamus are identified via known thalamo-cortical topographic segregation. While these studies report reasonable results for major nuclei, the low spatial resolution and signal-to-noise ratio (SNR) of clinical DTI and fMRI are limiting factors in differentiating subdivisions of those nuclei. High fiber orientation variability in DTI, induced by crossing fibers, further raises issues for segmentation of internal substructures. Results obtained

with different approaches may lack consistency and need further validation, which requires a ground truth [15].

Over the years, there has been an increasing interest in developing novel MR imaging protocols for direct visualization and anatomical delineation of the thalamic structures. Early work from Deoni *et al.* reported that T1 values ranging from 700 to 1400 ms at 1.5 T reveal different contrast in differentiating intrathalamic nuclei [16]. Through the use of different inversion times (TI), modifications of the magnetization-prepared rapidly-acquired gradient echo (MPRAGE) sequence have been proposed to visualize boundaries between some thalamic structures through intensity variation caused by difference in myelin concentration [17-18]. Susceptibility weighted imaging (SWI), on the other hand, is able to characterize iron deposition. Thanks to recent development in high-field imaging, the quality of such sequences acquired at lower field can be further improved with higher spatial resolution and SNR. Recently, to enhance the intrathalamic contrast, Tourdias *et al.* [19] presented a protocol to obtain an optimal MPRAGE sequence at 7T, while Abosch *et al.* [20] investigated the use of 7T SWI images. While finding the perfect sequence to visualize a specific structure is preferred in clinical settings because of limitation in scanning time, such constraint does not apply to atlas generation. In fact, as histological atlases require the use of multiple chemical markers to reveal different intrathalamic contrast, information across regions of different tissue properties can be enriched through different MR sequence parameters. Limiting acquisition to only one or one type of sequences to match the contrast of what has been done through multiple histological markers is challenging. In this chapter, building on previous work by Newton *et al.* [21], we follow a similar multi-contrast approach as used in histological delineation to design a method in which

we exploit the benefit of combining multiple high resolution MR sequences and a "frankenstein" approach to segment the internal structures of the thalamus.

Once the thalamus atlases are created from 7T images, the ultimate goal is to use them for automatic segmentation of thalamic structures in clinical T1-w images. Atlas-based techniques could be applied to project these structures onto the clinical 3T T1-w image. Single-atlas-based methods rely on registration to establish an accurate mapping within the thalamus region between the reference volume and the target. The accuracy of such method can be affected by the shape differences between the atlas and the target. To address this, multi-atlas-based methods fuse the labels transferred from several atlases in a statistically meaningful way. However, for homogeneous regions such as the interior of the thalamus, registration is mainly controlled by the smoothness of the deformation field. As a consequence, the shape of the intrathalamic structures may not be preserved after the non-rigid deformation and/or label fusion. These may lead to inaccurate localization of those nuclei and unrealistic shapes.

In contrast to atlas-based approaches, shape modeling has been applied to estimate structures with limited contrast by exploiting their relationship with regards to visible structures that are close. Originally proposed by Cootes *et al.* [22], statistical shape models can learn patterns of variability from a set of training shapes described by homologous landmark points. This approach has been used by many to solve a variety of segmentation problems. For instance, Blanc *et al.* predicted bone shape from surgically relevant predictors using statistical shape modeling and multivariate regression techniques [23]. Baka *et al.* incorporate landmark uncertainty into the shape regression to estimate the femur [24]. More recently, regression forests have been proposed to model the shape dependency between the subthalamic nucleus and its predictors extracted from clinical images [25]. In line with this principle, we hypothesize that

thalamic substructures could be inferred by their relative shape relationships with regard to the thalamus, which can be readily segmented in 3T volumes.

In this chapter, we first develop a protocol to generate human thalamus atlases using a series of 7T images. The second part of this chapter is dedicated to the use of such atlases for thalamic nuclei segmentation in clinical images. We describe atlas-based approaches using a single atlas and multiple atlases. We also propose shape-based approaches that build statistical shape models of all structures and predict the internal nuclei by fitting the thalamus to the models. Performance of these methods is compared in a leave-one-out validation of nine subjects.

6.2. Methods

In this section, we describe our protocol for generating the thalamus atlases and present our methods for segmenting the internal thalamic structures. We first discuss the generation of the atlases, the thalamic structures we have delineated, and the volumetric and shape analysis we have performed for those manual segmentations. Then in the following sections, we detail our approaches for segmenting intrathalamic structures with the 7T thalamus atlases using single atlas (SA), multiples atlases (MA), and statistical shape models (SSM).

6.2.1. Atlas Generation

To delineate thalamic nuclei, Tourdias *et al.* recently studied a range of TIs to design an MPRAGE sequence that provides the best intrathalamic contrast [19]. In our study, instead of limiting the acquisitions to one sequence, we acquired a set of MPRAGE sequences with TIs ranging from 400 ms to 1200 ms. Meanwhile, similar to Abosch *et al.* [20], we also acquired several high resolution SWI sequences in different orientations. While defining our protocol of

acquisition, one healthy subject (37 years old) was scanned during 4 sessions of 2 hours in a 7T Philips Achieva human whole body MRI scanner with a 32 channel receive-only head coil. This resulted in 25 sequences including MR T1-w, MPRAGE, T2-w, T2* and SWI images with full and partial head coverage. All sequences were first rigidly registered to a 7T full head T1-w image. These sequences were then loaded in a visualization and segmentation software to further characterize thalamic nuclei and delimit sub-territories of functional significance for stereotactic explorations. Switching between sequences revealed a sub-compartmentalization of thalamic nuclei into several groups.

From all sequences, a total of 7 sequences were found to be useful. Figure 6.1 shows a portion of one slice of those co-registered sequences in axial and coronal orientations overlaid with a manual thalamus surface. As illustrated by the yellow arrows, a wide range of contrast can be achieved from various sequences to assist in differentiating intrathalamic structures. While some of the sequences reveal similar intensity variation patterns that reinforce our confidence in depicting the internal boundaries, others provide complementary information that allows the delineation of different nuclei structures. Specifically, MPRAGE TI 400 to 1200 reveal the mammillothalamic tract, the mediodorsal nucleus, the parafascicular nucleus, the Habenulathe nucleus, the ventral anterior nucleus, the ventral lateral complex, and to a lesser degree also, lateral part of the centre median nucleus, inferior part of lateral posterior nucleus and lateral dorsal nucleus, lateral subdivisions of the pulvinar complex, and subdivisions of the ventral posterior complex. In SWI sequences, contrast is prevalent in the antero, ventral, and dorsal nucleus, medial nuclei group (mediodorsal nucleus, midline as a whole, and Habenula), and the pulvinar complex. The complementary distribution of contrasts among the MPRAGE and SWI

sequences appears to correlate with what has been illustrated by Morel *et al.* [5], which will be further illustrated in the result section.

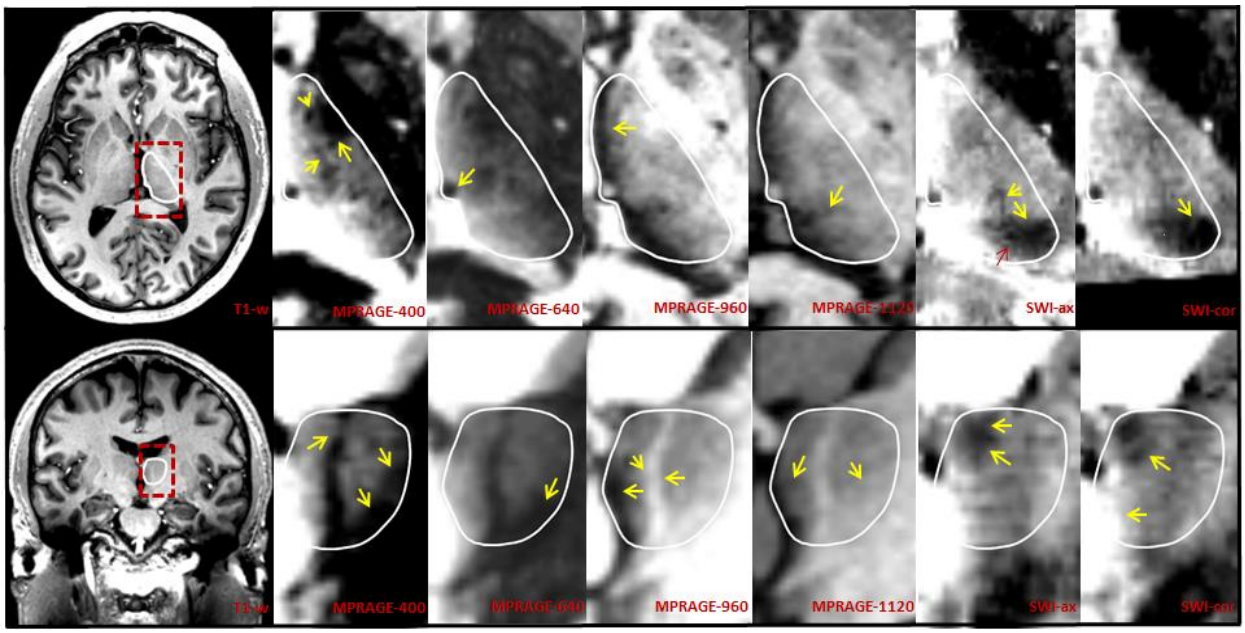

Figure 6.1. Representative slices of the 7T images acquired for one subject in axial (top row) and coronal (bottom row) views. For each row, the leftmost panel is the 7T T1-w image of the whole brain. The right panels show a series of 7T images for the thalamus region, which are MPRAGE-400 (TI = 400), MPRAGE-640 (TI = 640), MPRAGE-960 (TI = 960), MPRAGE-1120 (TI = 1120), SWI-ax (acquired in axial direction), and SWI-cor (acquired in coronal direction) respectively. For each sequence, the yellow arrows point to the plausible intensity variations inside the white thalamus contours.

To perform manual delineation, the co-aligned 7T sequences were normalized into a stereotactic coordinate system as described in Chapter III by applying a rigid transformation matrix estimated from manually selected Anterior Commissure, Posterior Commissure, and midsagittal plane. The thalamus and a total of 23 intrathalamic substructures were manually delineated and validated by an expert for both hemispheres. Their names and their spatial hierarchy as suggested by Morel *et al.* [5] are shown in Figure 6.2. While we are able to visualize most of these structures, contrast does not allow us to segment some less visible or more complex structures. For example, the anteromedial nucleus, division of the midline nuclei, and

the boundary between the lateral dorsal nucleus and lateral posterior nucleus cannot be easily differentiated. As needed for our study, we used a-priori knowledge to assist in manual delineation. Such a-priori information consists of knowledge of the relative position and shape of the nuclei to be segmented based on Morel histologic thalamic parcellation [5].

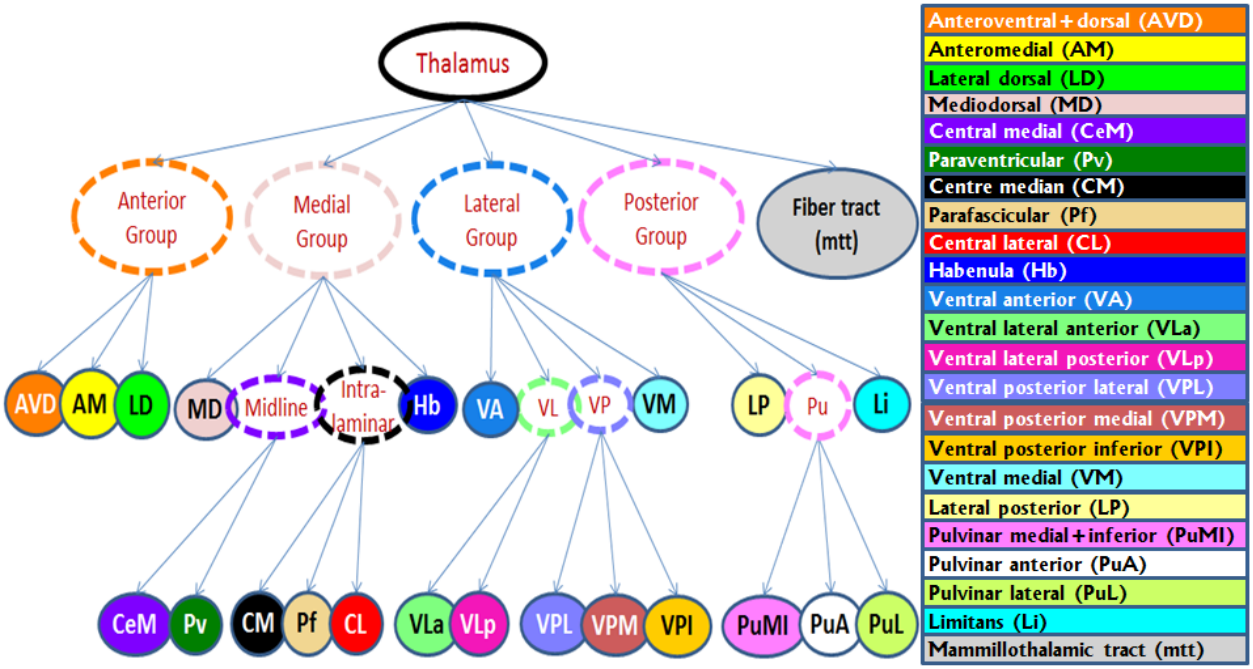

Figure 6.2. Hierarchical relationship of the thalamus and the thalamic nuclei involved in this study using the Morel nomenclature.

For this study, the above process was reproduced on eight healthy subjects (seven male, two female). This includes acquiring seven imaging sequences at 7T (a standard T1-w image, four MPRAGE sequences, and two high in-plane resolution SWIs) and manually delineating the described thalamic structures for both hemispheres. The T1-w and MPRAGE sequences were acquired with varying inversion time values with an inversion prepulse (shot interval = 4500ms). The SWIs were acquired in two different directions using flow-compensated slice selective gradient echo. More details about the acquisition parameters are presented in Table 6.1.

Table 6.1. Imaging parameters for each sequence at 7T used in this study

	Direction	FOV (mm)	Dimension ($voxel^3$)	Resolution (mm^3)	TR/TE (ms)	TI (ms)	Flip Angle (˚)	Acquisition Time (min)
T1-w	Axial	246×246×245	352×352×350	0.699×0.699×0.7	4.74/2.1	0	7˚	8.7
MPRAGEs	Axial	246×246×174.3	352×352×249	0.699×0.699×0.7	4.74/2.1	400 640 960 1120	7˚	8.7
SWIs	Axial Coronal	240×240	1024×1024×50	0.234×0.234×1.1	1676/19.49	-	45˚	10.6

FOV: field of view; TR/TE: repetition time/echo time (in milliseconds); TI: inversion time (in milliseconds).

Together with the 7T acquisition, a 3D T1-w MR volume was also acquired for each subject with a 3T Philips Achieva scanner using the following parameters: FOV = 256×256×170 mm; resolution = 1 mm isotropic; TR = 13.0ms; TE = 3.65ms; nominal flip angle = 8˚; acquisition time = ~13 minutes. This 3T T1-w volume serves as the target image for semi-automatic segmentation of thalamic nuclei.

6.2.2. Analysis of Manual Segmentations

Manual delineations of the structures using the above protocol result in surfaces represented by triangulated meshes. To measure their volumes, each surface is converted into a binary mask. Volumes of these structures are compared across subjects for each side of the brain as well as for the two sides pooled together.

To study the shape of each substructure, we use a point distribution model to represent each structure with a set of landmark points. Structures on the left side of the brain are flipped to the right side along the mid-sagittal plane and treated as different instances. For each structure, correspondence of those landmarks between difference instances is established using a series of registration. Surfaces of the same structure are then aligned to calculate the mean shape. Details of this process will be presented in section 6.2.3.3.

6.2.3. Semi-automatic Segmentation of Thalamic Nuclei

With a number of thalamus atlases created from high resolution 7T images, our task is to segment the thalamic nuclei in clinically available 3T T1-w images, in which these nuclei are not visible. In this work, we assume that the thalamus as a whole has been segmented, and this could be achieved with various existing tools such as FreeSurfer [26]. Given the thalamus as input, the subsections below demonstrate the steps that are used to estimate the internal nuclei in 3T T1-w images using various approaches.

6.2.3.1. Single-atlas-based Segmentation

We randomly choose a subject out of the nine volunteers as the reference. Segmentation using this approach relies on accurate mapping between this reference volume and a target volume. To achieve this, a series of registrations are performed to best align the thalamic region of the two volumes. This includes a three-step image registrations from the target 3T T1-w image to the atlas 7T T1-w image, i.e., 1) a global affine registration using an intensity-based technique with mutual information as a similarity measure [27-28], 2) a local affine registration which restricts the computation of the mutual information to the deep brain region, and 3) a nonrigid registration using the Adaptive Bases Algorithm [29]. Thalamic structures in the atlas, represented as triangulated meshes, can be projected onto the target space using the inverses transformations.

With a segmentation of the target thalamus, we further refine the alignment of the thalamic region by non-rigidly registering the thalamus surface projected from the atlas to the target thalamus surface. This is done with an automated algorithm developed in house. This algorithm acts as 3D snake [30] and displaces each vertex p_i in source surface S to match the target surface T. The energy functional it tries to minimize is the sum of vertex energies:

$$E(S') = \sum_{i=1}^{n} E(p_i') = \sum_{i=1}^{n} (E_{distance}(p_i', T) + \alpha \cdot E_{smooth}(p_i') + \beta \cdot E_{curv}(p_i', S)) \quad (6.1)$$

where the weights α and β balance between the external energy and the two internal energy terms. The external energy $E_{distance}$ minimizes the distance between each deformed point p_i' and the target surface T, resulting in a closest point matching alike algorithm. The first internal energy E_{smooth} performs local surface smoothing by making the vertex stay closer to the centroid of its adjacent vertices, while the second one E_{curv} maintains the topology of the surface by keeping the difference between each vertex and its adjacent vertices as similar as possible to the one in the original surface S:

$$E_{smooth}(p_i') = \left\| p_i' - \frac{1}{|A_i|} \sum_{j \in A_i} p_j' \right\|^2 \quad (6.2)$$

$$E_{curv}(p_i', S) = \frac{1}{|A_i|} \sum_{j \in A_i} \left\| (p_i' - p_j') - (p_i - p_j) \right\|^2 \quad (6.3)$$

where A_i represents the set of vertices adjacent to vertex p_i. By iteratively updating S' to minimize (6.1), we obtain an optimal surface S^* deformed from S that preserves the smoothness and curvature but also well aligned with the target surface T.

As S and S^* form two sets of corresponding points, we can further approximate a deformation field from the above surface-based registration using thin-plate-splines (TPS). As performed in [31-32], the TPS transformation minimizes:

$$E(T_{TPS}) = \sum_{i=1}^{n} \|p_i^* - p_i\|^2 + \lambda \cdot \sum_{\alpha_1+\alpha_2+\alpha_3=2} \frac{2}{\alpha_1!\,\alpha_2!\,\alpha_3!} \iint \left[\left(\frac{\partial^2 T_{TPS}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \partial x_3^{\alpha_3}} \right)^2 \right] d\vec{x} \quad (6.4)$$

where the first term of the functional measures the difference between each original point p_i in S and its deformed position p_i^*, and the second term, weighted by λ, controls the smoothness of deformation field.

To segment internal thalamic substructures, we apply the transformations estimated from the image-based registrations as well as the TPS transformation estimated from the surface-based registration in reverse orders. This process is illustrated in Figure 6.3.

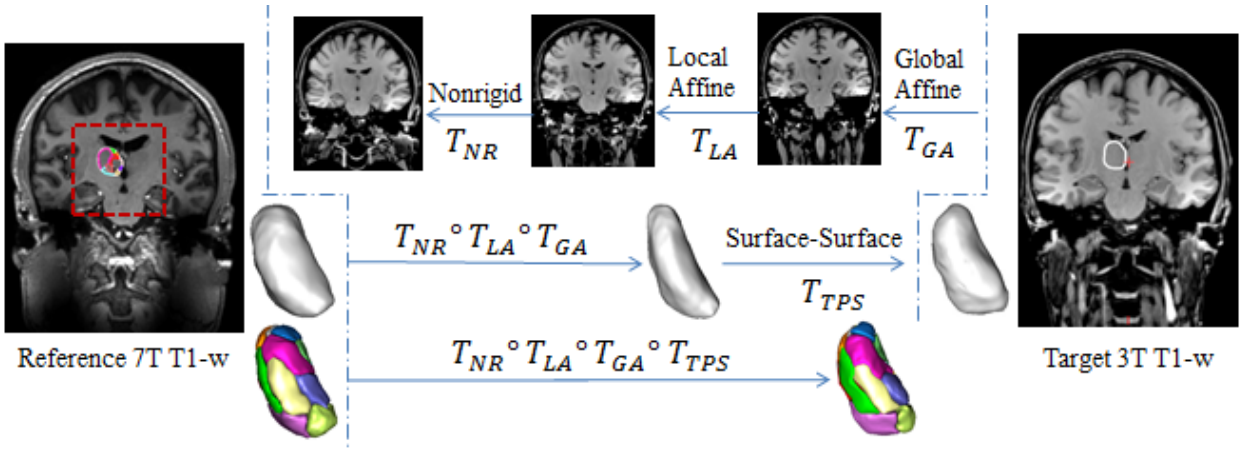

Figure 6.3. Flowchart of thalamic nuclei segmentation using a single atlas with a given thalamus.

6.2.3.2. Multi-atlas-based Segmentation

Segmentation using a single atlas may be biased by the anatomy of the selected reference volume. This can be avoided by using several subjects as references, as performed in multi-atlas-based approaches. In this chapter, we treat all the subjects except for the target as atlases and register them to the target following the single-atlas-based segmentation process. This produces multiple segmentations per structure from each of the atlases. To fuse those segmentations, we represent results obtained with each atlas as a discrete mask and use majority voting to assign labels to each voxel, i.e., the label at a voxel is determined by the maximum occurrence of labels from all segmentations.

6.2.3.3. Statistical Shape-based Segmentation

The position of thalamic nuclei may also be inferred from their relative location and shape relationship with the thalamus. We use statistical shape models to do this, which requires localizing homologous points across shapes, computing the eigenvectors of the covariance matrix, and fitting the thalamus to the shape models.

Establishing Point Correspondence

To establish point correspondence between training shapes, we select a reference shape in the same reference volume as used in section 6.2.3.1 but on the right side of the brain. Structures on the left side and/or of another subject are mapped onto the reference shapes.

Correspondence for the thalamus is established using image-based registration and surface-based registration, similarly to the steps performed in Figure 6.3 but using 7T T1-w images of the target. Note that for shapes on the left hemispheres, the target image is flipped in the lateral direction in the global affine registration step to map to the right side. For the internal thalamic structures, surfaces are first transformed from the atlas to the subject as done in Figure 6.3. We then nonrigidly register the transformed surface to the target surface using the algorithm described in section 6.2.3.1. Note that this algorithm performs well if the two surfaces are initially close. If substantial differences between two shapes exist, as is sometimes the case for nuclei substructures, this algorithm may fail. When this happens, we manually adjust the transformed surface to bring it closer to the target surface, and rerun the algorithm. By repeating this process, correspondence for all training structures is established.

Determining Joint Relationship

To model the relationship between the thalamus and the nuclei, we build a large shape model that contains the thalamus and all the 23 nuclei. This is the straightforward approach to infer the shapes of internal structures; we call this approach non-hierarchical SSM (NHSSM).

However, it is much more complex for a single model to capture the shape variations and relationships of so many structures, especially with limited training data. Therefore, instead of jointly modeling all the structures we further propose to hierarchically model the shapes of those structures following the decomposition shown in Figure 6.2. Starting at the top level of the hierarchy, we first jointly model the thalamus and the general nuclei groups. At the next level, each group is jointly modeled with its subdivisions, and the process is repeated until the leaf nodes are reached. The individual shapes at each hierarchy level are shown in Figure 6.4. In this way, the top level models capture a more global view of those internal structures, whereas the complexity of the shapes of individual nuclei or sub-nuclei is delegated into lower level models. We call this approach hierarchical SSM (HSSM).

To group the nuclei as illustrated in Figure 6.4, we combine the surfaces of the nuclei in this group into one mesh by removing the inner boundaries where two nuclei are adjacent to each other. Identification of vertices that belong to inner boundaries is done on the reference subject by computing the distance from each vertex to each other surface based on its distance map obtained by a fast marching method [33]. Once such vertices are identified, this is subsequently propagated to the rest of the subjects via the established vertex correspondence.

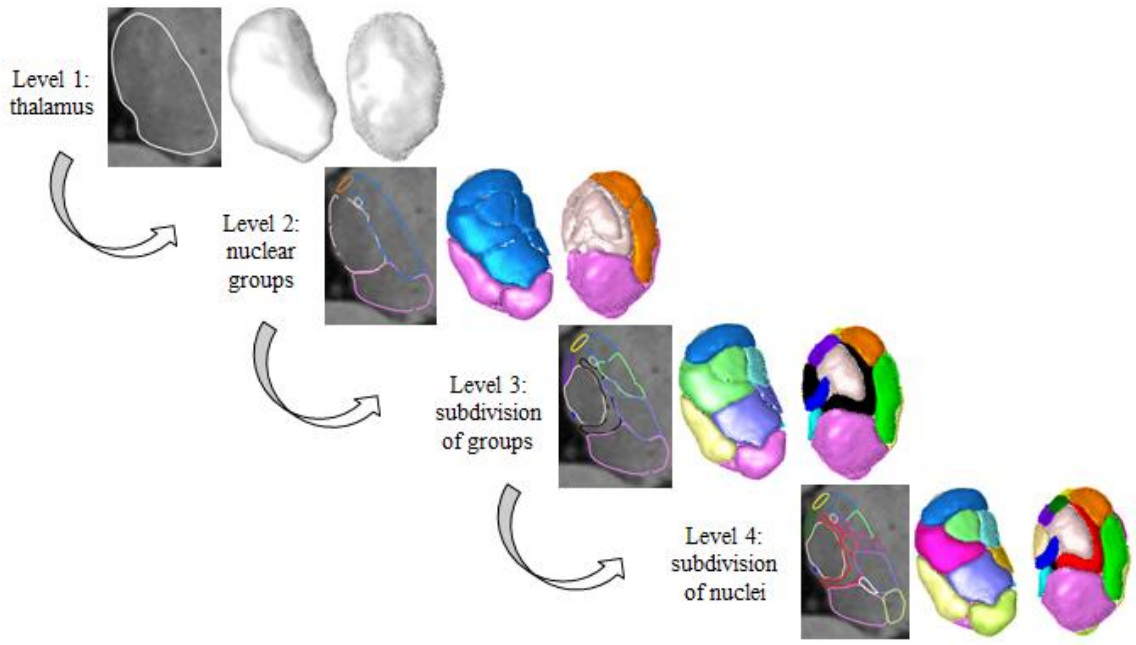

Figure 6.4. Individual shapes used for modeling at each level. For each the left panel is a coronal slice overlaid with 2D contours of the shapes, and the right panels are 3D views in two different orientations.

Constructing Statistical Shape Models

Once correspondence is established and joint relationships are determined, training surfaces can be registered to the reference surfaces with 7-DOF transformation (three rotations, three translations, one isotropic scaling) using a standard Procrustes approach [34]. When HSSM is used, this is performed separately for individual joint models, i.e., only the surfaces involved in this model are co-registered to build this particular model. This allows measuring the anatomical variations of thalamic structures at different scales using different models across the training set. To build the shape model, we perform the following procedure as described by Cootes *et al.* [22]. First a mean shape of the co-registered corresponding point set is extracted and the covariance matrix is computed as the deviation of the point set from the mean shape. The eigenvectors of the covariance matrix represent the modes of variation and their eigenvalues the explained variance in the training set. A new shape can be approximated by adding a linearly

weighted sum of the eigenvectors within three standard deviations of the mean to the mean shape:

$$x = \bar{x} + Pb \tag{6.5}$$

where $\bar{x}$ is the mean shape, P is a set of modes of variation, and b is the parameters that determines the new shape x.

Segmenting Internal Structures via Hierarchical Weighted Fitting

Given a thalamus surface in a new 3T volume, to use our models to segment the thalamic nuclei, we first establish a correspondence between the thalamus surfaces as we have done for the single-atlas-based segmentation. For the NHSSM approach, we simply fit the joint model directly to the thalamus. For the HSSM approach, a hierarchical shape fitting algorithm is applied to localize each nucleus. We start by fitting the joint model at the top level to the thalamus. This joint shape includes the thalamus and all general nuclei groups. The fitting to this joint model is done by weighting the thalamus surface points as 1 and the rest as 0. Neglecting the alignment, the shape x_{new} estimated in a weighted fashion given the current joint shape x_{cur} is obtained by:

$$x_{new} = \bar{x} + Pb, b = (P^T W^T W P)^{-1} P^T W^T W \cdot (x_{cur} - \bar{x}) \tag{6.6}$$

where W is a diagonal matrix formed by weights for each surface point, with value 1 for thalamus and 0 for the rest. This permits us to infer the position of the general nuclei groups. Subsequently, given the external boundary of the nuclei group estimated at the previous level, the internal boundary between the subdivisions of the group is inferred by fitting the external boundary to the corresponding joint shape model.

After the shape fitting we have observed that the external boundary of the fitted thalamus did not exactly fall on the boundary of the segmented thalamus. We attribute this small fitting

error to the small size of our training set. To correct for this fitting error, a TPS transformation as described in section 6.2.3.1 is estimated between the original thalamus boundary and the fitted boundary and then applied to all the boundary points localized by the model. This correction is performed after each shape fitting step, i.e., at the last step of NHSSM and at each hierarchical level for HSSM.

6.2.4. Validation of Semi-automatic Segmentations

To evaluate the performance of the above methods, we perform segmentations in 3T T1-w images of each subject in a leave-one-out fashion, i.e., the volume being segmented is left out of the model construction. The subject that serves as the reference for point correspondence is excluded from testing, which leaves us sixteen test cases (two hemispheres of eight subjects). As mentioned earlier, we assume the thalamus has been well segmented, and use its manual segmentation as an input to feed our algorithm. This requires each 7T T1-w image to be rigidly registered to the corresponding 3T T1-w image and the manual structures delineated in 7T space to be projected to the 3T volume. Accuracy of the results is measured by dice coefficients and mean surface errors in the 3T space compared to the manual segmentations after projection.

6.3. Results

6.3.1. Manual Segmentations

Thalamic structures obtained following the atlas generation protocol are shown in representative slices in axial, coronal, and sagittal views in Figure 6.5, 6.6, and 6.7 respectively. For each orientation, two image sequences of the same slice are selected to illustrate the complementary contrast for manually delineating different intrathalamic structures. The top rows of these three

figures present slices of the same subject, whereas the bottom rows are of different subjects. When comparing the two rows, we observe a similar range of contrast and ability to separate substructures. This indicates the consistency of the acquired sequences across individuals. In each figure, the bottom right panel displays the 2D histological drawing that corresponds to the selected slices of the two subjects. Compared to this panel, manual contours derived from intensity variation and spatial priors in 7T sequences correspond well with known anatomical structures of the thalamus from histology. This suggests the feasibility of using such sequences to generate thalamus atlases.

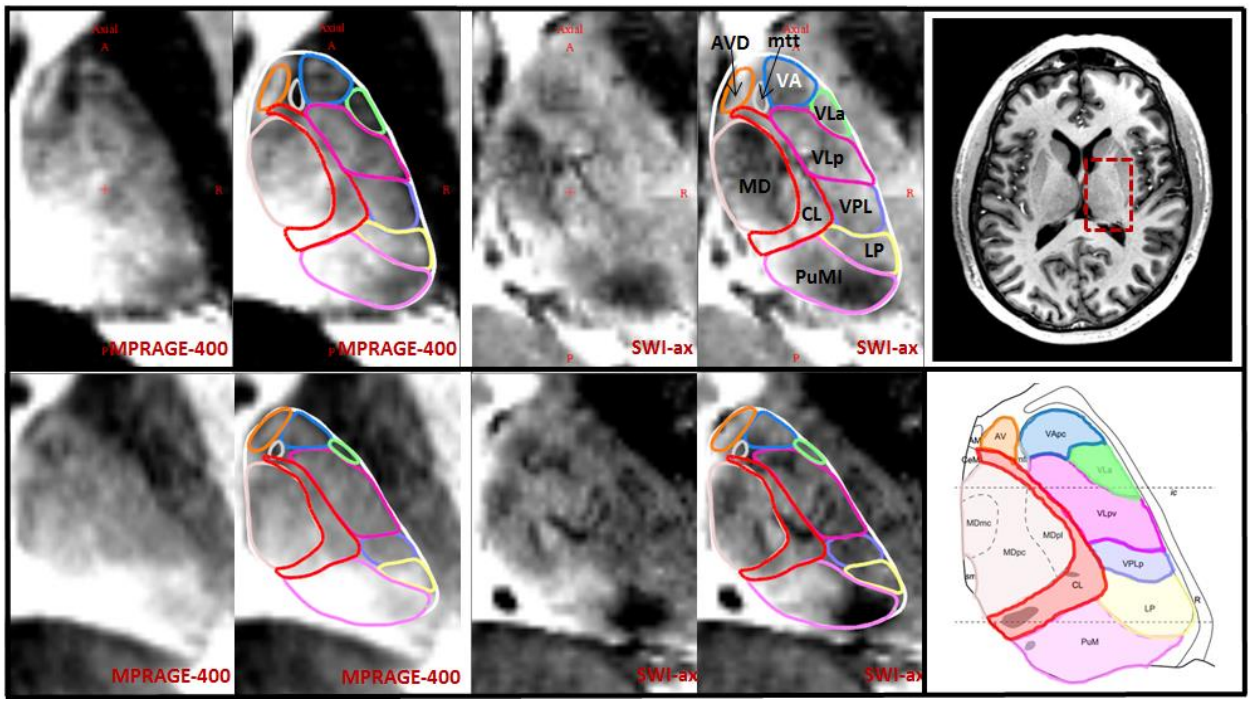

Figure 6.5. Representative axial slices of two subjects in the top and bottom rows respectively. From left to right, each row presents a portion of the MPRAGE-400 and SWI-ax sequences in the thalamus region, together with manual segmentation of thalamic structures overlaid on top. The bottom right panel shows the corresponding axial histological drawing in Morel atlas [5].

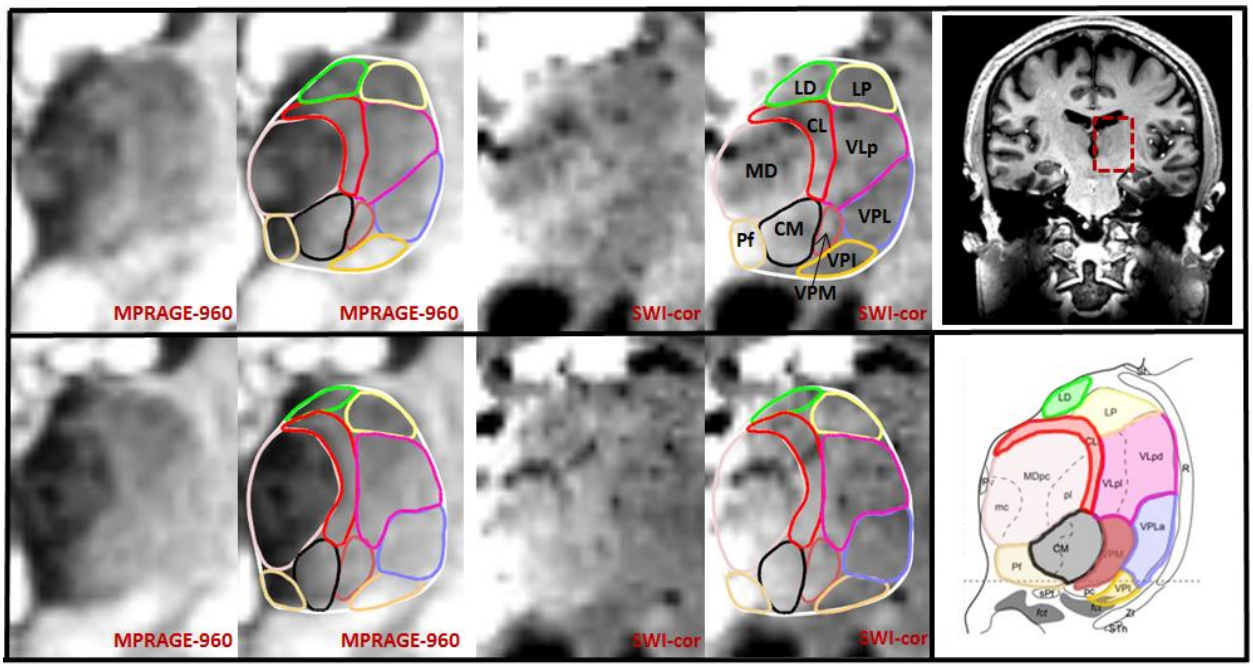

Figure 6.6. Representative coronal slices of two subjects in the top and bottom rows respectively. From left to right, each row presents a portion of the MPRAGE-960 and SWI-cor sequences in the thalamus region, together with manual segmentation of thalamic structures overlaid on top. The bottom right panel shows the corresponding coronal histological drawing in Morel atlas [5].

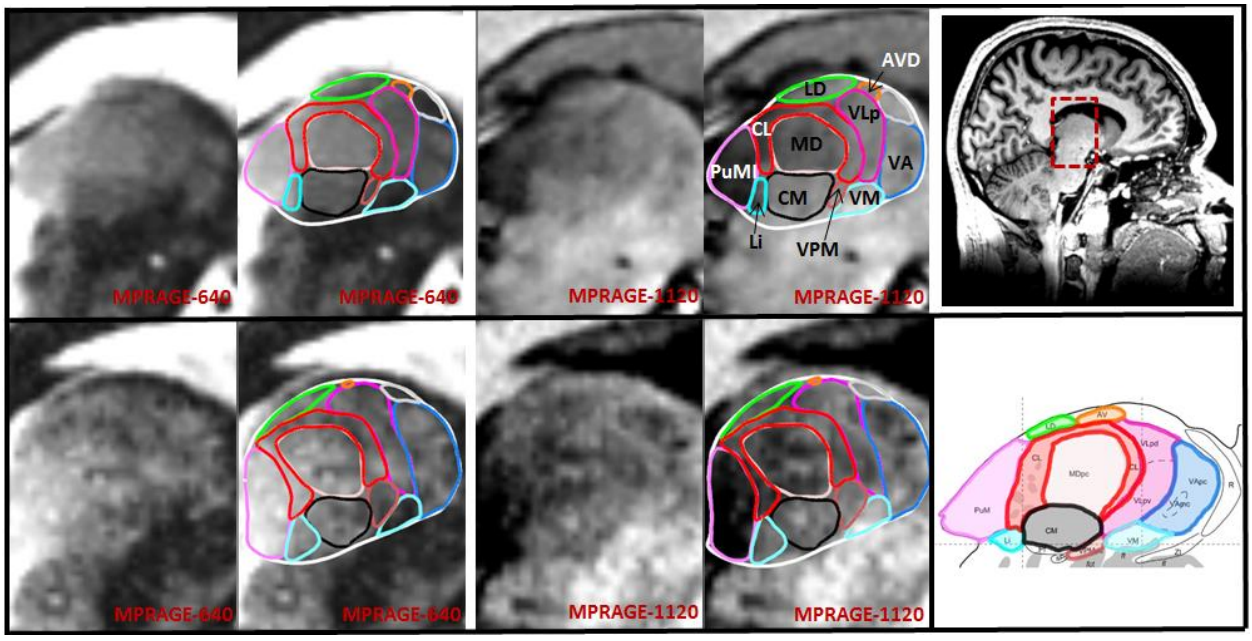

Figure 6.7. Representative sagittal slices of two subjects in the top and bottom rows respectively. From left to right, each row presents a portion of the MPRAGE-640 and MPRAGE-1120 sequences in the thalamus region, together with manual segmentation of thalamic structures overlaid on top. The bottom right panel shows the corresponding sagittal histological drawing in Morel atlas [5].

The nine subjects involved in this study have an average inter-commissural distance of 24.8 mm ranging from 21.6 mm to 28.3 mm. Table 6.2 reports the volumes of each structure for the nine subjects on the left and right side of the brain. Mean, standard deviation, and coefficient of variation of volumes are calculated per structure for each side as well as for the two hemispheres pooled together and are also shown in this table. The volume of the thalamus varies from 4659 mm^3 to 8187 mm^3, which mainly affects large substructures such as MD, CL, VA, VLp, VPL, and PuMI. Coefficients of variation for AM, Hb, PuA, and Li are above 40%, a relatively large number compared to those of other structures. This may be explained by the substantial variation in sizes for those structures across subjects and/or the confounding errors in manual delineation that are difficult to measure.

Table 6.2. Volumes (in mm^3) and their statistics of each structure for the nine subjects

Structures	Side	Subject 1-9									Mean		Std.		CV	
THAL	Left	6916	7963	7106	6468	8113	5666	7171	5058	6563	6763.2	6771.8	1085.1	1006.6	16.0%	14.9%
	Right	6577	7814	7383	6794	8187	5634	6664	4659	7157	6780.4		987.8		14.6%	
AVD	Left	203	129	103	169	139	136	173	146	154	142.4	146.2	25	26.5	17.5%	18.1%
	Right	127	135	108	165	162	178	120	124	163	150		29		19.3%	
AM	Left	24	36	34	27	16	8	33	6	31	30.2	27.1	25.9	19.6	85.8%	72.5%
	Right	25	24	24	15	94	11	26	10	41	23.9		11.2		46.8%	
LD	Left	198	139	106	104	235	122	116	118	137	139.7	140.7	52.6	47.6	37.6%	33.8%
	Right	141	99	118	109	253	200	105	126	107	141.7		45.2		31.9%	
MD	Left	674	867	814	702	847	549	818	522	660	683.5	700.3	137.2	129.7	20.1%	18.5%
	Right	686	854	757	796	841	520	581	489	626	717.1		127.7		17.8%	
CeM	Left	31	33	40	17	32	22	27	41	34	27.2	29.1	7.3	7.6	26.7%	26.3%
	Right	25	35	32	13	24	22	31	29	36	30.9		8		25.9%	
Pv	Left	28	22	24	22	24	9	39	18	24	22.4	22.8	8.5	8.0	37.9%	35.1%
	Right	16	35	23	23	18	7	19	32	28	23.3		8		34.3%	
CM	Left	228	221	215	128	236	102	190	86	188	153.1	165.2	52.2	54.6	34.1%	33.0%
	Right	208	226	173	120	192	71	121	103	164	177.2		57.3		32.3%	
Pf	Left	76	71	103	53	60	51	69	59	68	65	66.4	23.2	19.3	35.7%	29.0%
	Right	69	58	119	49	50	42	69	53	76	67.8		15.7		23.1%	
CL	Left	408	569	461	470	413	301	580	332	403	438.9	438.2	73.3	82.0	16.7%	18.7%
	Right	353	494	415	475	493	391	510	313	506	437.6		94.5		21.6%	

Hb	Left	31	8	10	11	18	14	13	18	5	11.6	12.9	5.1	6.5	44.2%	50.0%
	Right	22	9	5	15	15	13	9	9	8	14.3		7.6		53.5%	
VA	Left	315	388	296	205	335	92	410	192	259	231.9	254.3	77	90.2	33.2%	35.5%
	Right	217	280	222	171	287	74	317	213	306	276.8		101.2		36.6%	
VLa	Left	111	126	148	80	95	34	73	65	104	88.6	90.8	33.3	32.9	37.6%	36.2%
	Right	100	101	144	110	75	44	61	50	112	93		34.3		36.9%	
VLp	Left	740	996	725	858	953	667	888	804	717	784.2	800.3	157.6	134.6	20.1%	16.8%
	Right	635	879	947	743	971	600	914	555	814	816.5		114.5		14.0%	
VPL	Left	642	820	754	481	580	318	400	431	452	569	555.5	195.3	177.6	34.3%	32.0%
	Right	774	674	822	539	585	300	341	368	718	542		168.8		31.1%	
VPM	Left	86	95	83	73	76	40	59	52	74	75.7	73.2	29.6	23.7	39.1%	32.4%
	Right	127	110	75	72	60	29	71	52	86	70.8		17.4		24.6%	
VPI	Left	58	68	64	34	42	29	31	51	82	50.7	50.9	19.3	18.4	38.0%	36.1%
	Right	80	81	52	39	47	33	26	42	56	51		18.6		36.4%	
VM	Left	107	114	82	70	52	47	79	63	93	77.7	78.1	24.1	22.9	31.0%	29.3%
	Right	98	96	90	96	54	44	61	54	107	78.5		23.1		29.4%	
LP	Left	364	225	274	308	348	359	239	220	257	298.3	293.3	73.8	64.6	24.7%	22.0%
	Right	419	247	283	259	393	366	242	220	257	288.3		58		20.1%	
PuMI	Left	906	1105	971	982	1501	1038	1068	720	1081	1223.7	1132.5	336	287.0	27.5%	25.3%
	Right	890	1498	1111	1317	1847	1126	1331	690	1204	1041.3		208.5		20.0%	
PuA	Left	127	266	108	191	119	105	120	78	181	132.6	138.3	66.2	60.8	49.9%	43.9%
	Right	189	280	103	138	111	87	88	64	132	144.1		58.3		40.4%	
PuL	Left	362	302	311	412	316	580	250	259	195	278.2	305.0	120.9	116.6	43.5%	38.2%
	Right	345	137	239	348	164	514	193	210	352	331.8		112.4		33.9%	
Li	Left	37	18	17	20	21	4	31	15	61	17.8	21.3	7.1	12.7	40.1%	59.9%
	Right	19	17	16	14	13	11	13	21	35	24.7		16.3		66.0%	
mtt	Left	54	89	56	50	75	28	53	21	61	45.1	49.6	10.1	16.7	22.4%	33.6%
	Right	45	58	54	55	38	39	29	37	52	54.1		21		38.9%	

For each structure, the mean, standard deviation (Std.), and coefficient of variation (CV) of volumes are reported, first for each hemisphere and then for the two hemispheres pooled together.

Figure 6.8 presents the manual substructures for the nine subjects. For each subject, 3D views in two different orientations are shown for structures on both left and right brains. For each structure of those subjects, its mean shape is shown in Figure 6.9. The shape of each nucleus is plotted together with a semi-transparent thalamus surface to illustrate its size and location with respect to the thalamus.

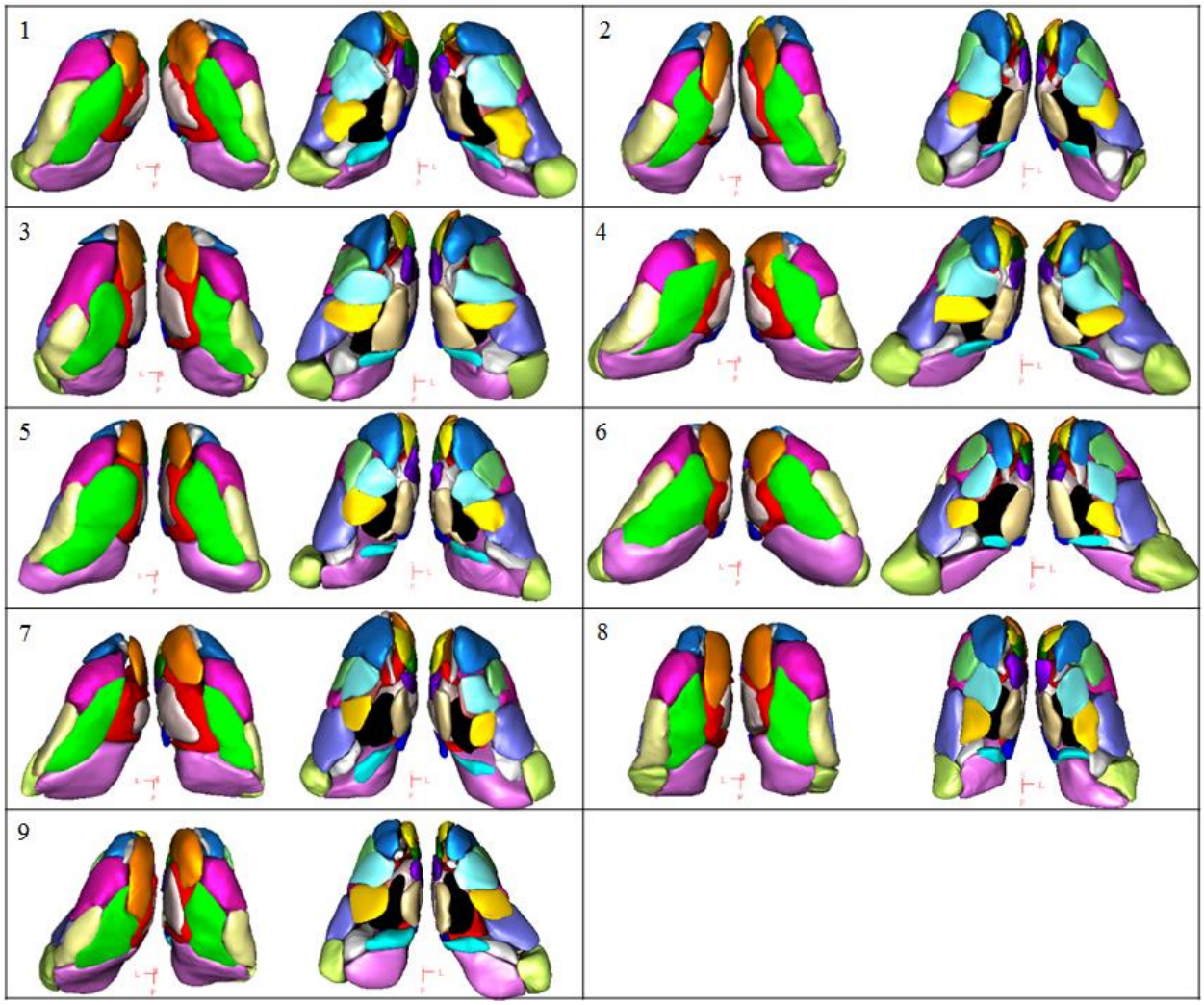

Figure 6.8. 3D views of all manual substructures for the nine subjects of both hemispheres.

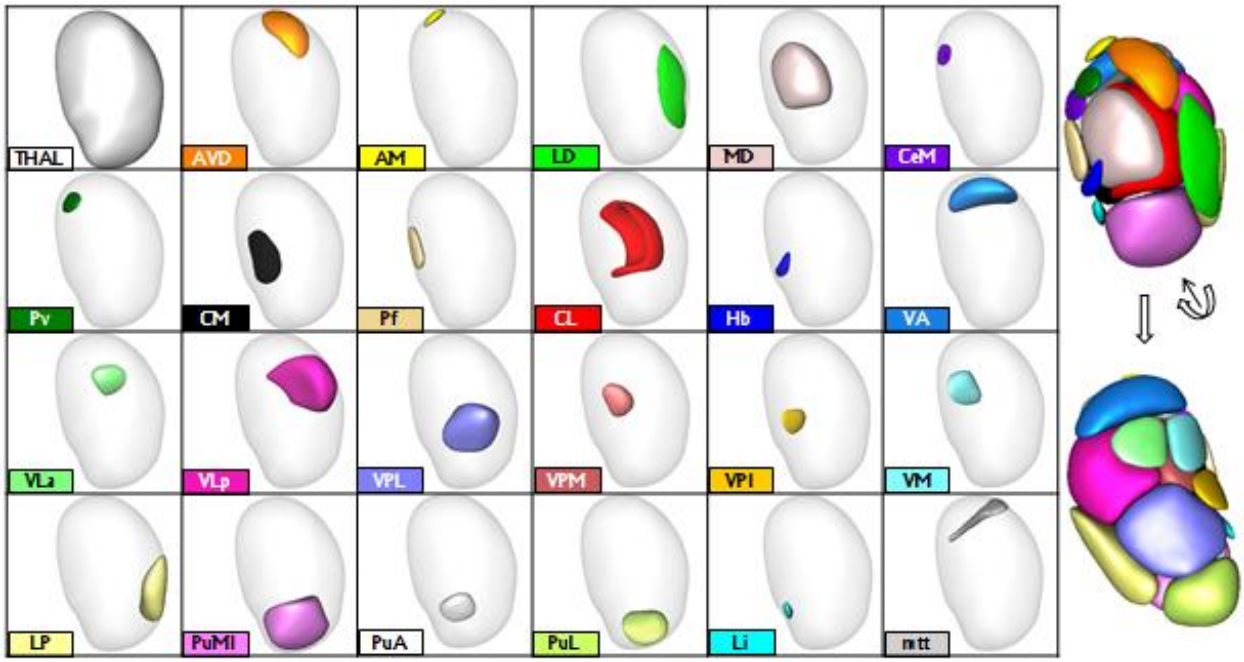

Figure 6.9. Mean shape of each structure for the nine subjects pooled from both hemispheres.

6.3.2. Semi-automatic Segmentations

In the leave-one-out validation study, segmentation results for one subject using the proposed methods are illustrated in Figure 6.10. As shown in this figure, segmentations obtained with different methods appear similar to each other, and they all agree relatively well with the manual delineations. Observation of the dark yellow VPI structure in the 3D views indicates that segmentations obtained by both the NHSSM and HSSM approaches tend to preserve the original shapes better than the ones using SA and MA approaches.

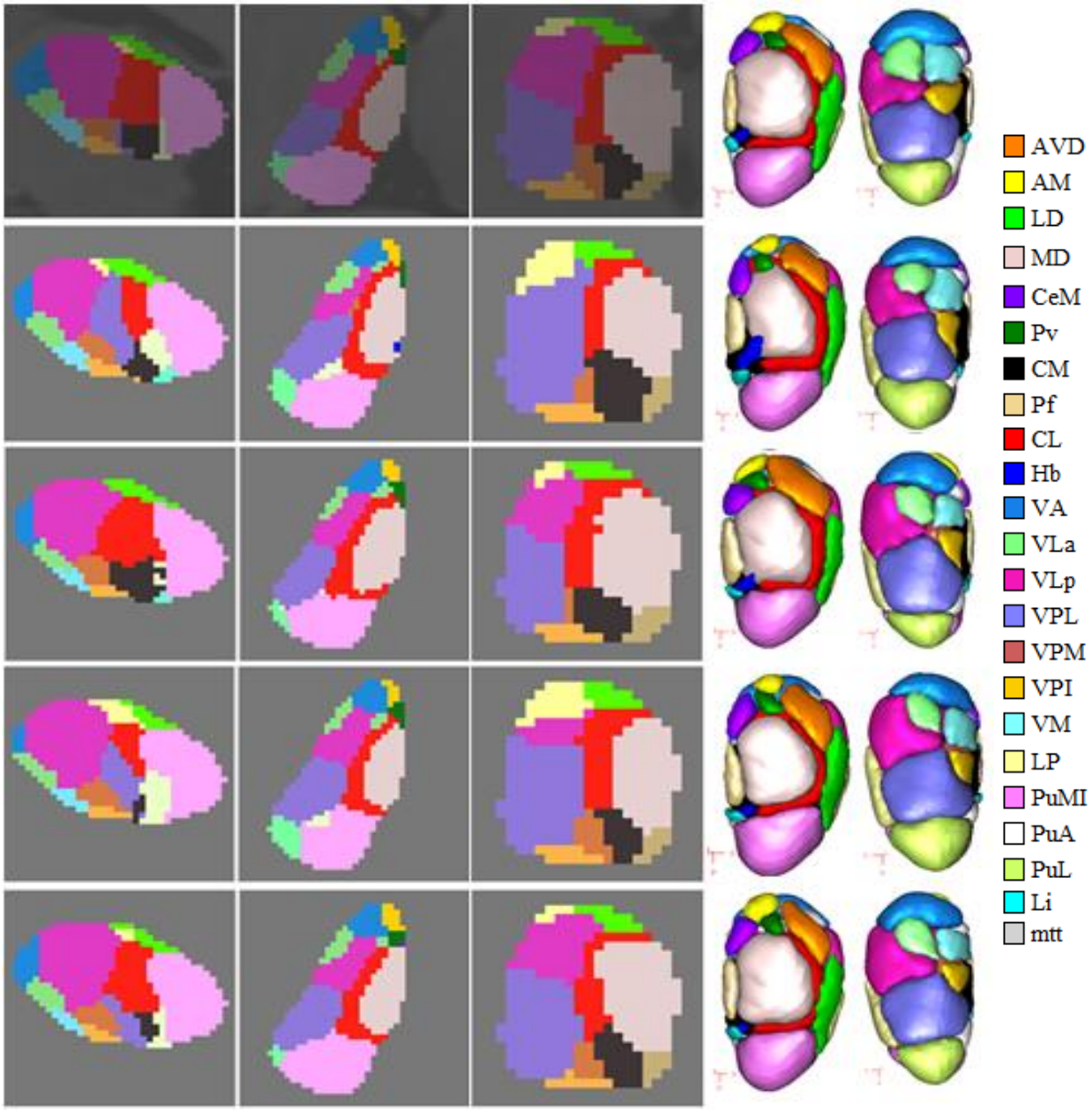

Figure 6.10. Qualitative visualization of the thalamic nuclei segmentation for one subject in the leave-one-out study. Top row illustrates the ground truth overlaid on top of the 3T T1-w image, and bottom rows are the results obtained by SA, MA, NHSSM, and HSSM approaches respectively. In each row, a sagittal, coronal, and axial slice are shown on the left, together with 3D views in two orientations on the right. Surfaces are smoothed for visualization purposes.

Figure 6.11 and 6.12 show the segmentation accuracy for each nucleus using the four methods measured in dice coefficients and mean surface errors (MSE) (in millimeter)

respectively. Their mean and the standard deviation are reported in Table 6.3. Results of the substructures are arranged in ascending order according to the mean volumes of manual delineations, with the smallest structure shown first. For those structures, mean dice values range from 0.25 to 0.80 using SA, 0.43 to 0.85 using MA, 0.46 to 0.87 using NHSSM, and 0.53 to 0.90 using HSSM; mean MSE values range from 0.34 mm to 1.34 mm using SA, 0.34 mm to 0.88 mm using MA, 0.27mm to 0.79 using NHSSM, and 0.27 mm to 0.64 mm using HSSM. Compared to SA, the use of multiple atlases leads to more robust segmentations. Results of paired Wilcoxon signed rank tests suggest that the differences in dice values between two atlas-based methods are statistically significant at significance level of 0.05 for each of the nuclei except for Hb, CeM, VPI, VM, VLa, LD, LP, and PuL, whereas the MSEs of these two methods statistically differ excluding CeM, mtt, VM, VLa, LD, AVD, VLp, and PuMI. Compared to SA and MA, the shape models consistently produce more accurate segmentations. The hierarchical shapes models perform better than the one without hierarchy in most cases, especially for structures of larger sizes such as VLp and PuMI. Results of the one-way ANOVA tests demonstrate that the group means of the four methods are significantly different for each nucleus at 0.05 significance level, except for VM, VLa, and PuL measured in dice and LD, AV, and PuL measured in MSE. We also observe from the figures that the size of the structures is correlated with the dice coefficient, whereas it is relatively independent for the MSE. For example, dice values of tiny structures such as Hb, Li, and Pv (10 to 30 voxel^3) are small and highly variable, but their MSEs remain relatively similar to those of other structures. This is expected, as the dice coefficient is sensitive to the size and the shape of the structures; a small deviation from the ground truth for a small-sized or elongated structure could lead to a large decrease in dice values. Nevertheless, these are encouraging results for structures with sizes ranging from 12.9 mm^3 (Hb)

to 1132.5 mm^3 (PuMI), especially for AVD, VA, LP, PuL, VPL, MD, VLp, and PuMI, with mean dice values above 0.8 and mean MSEs below 0.5mm using HSSM.

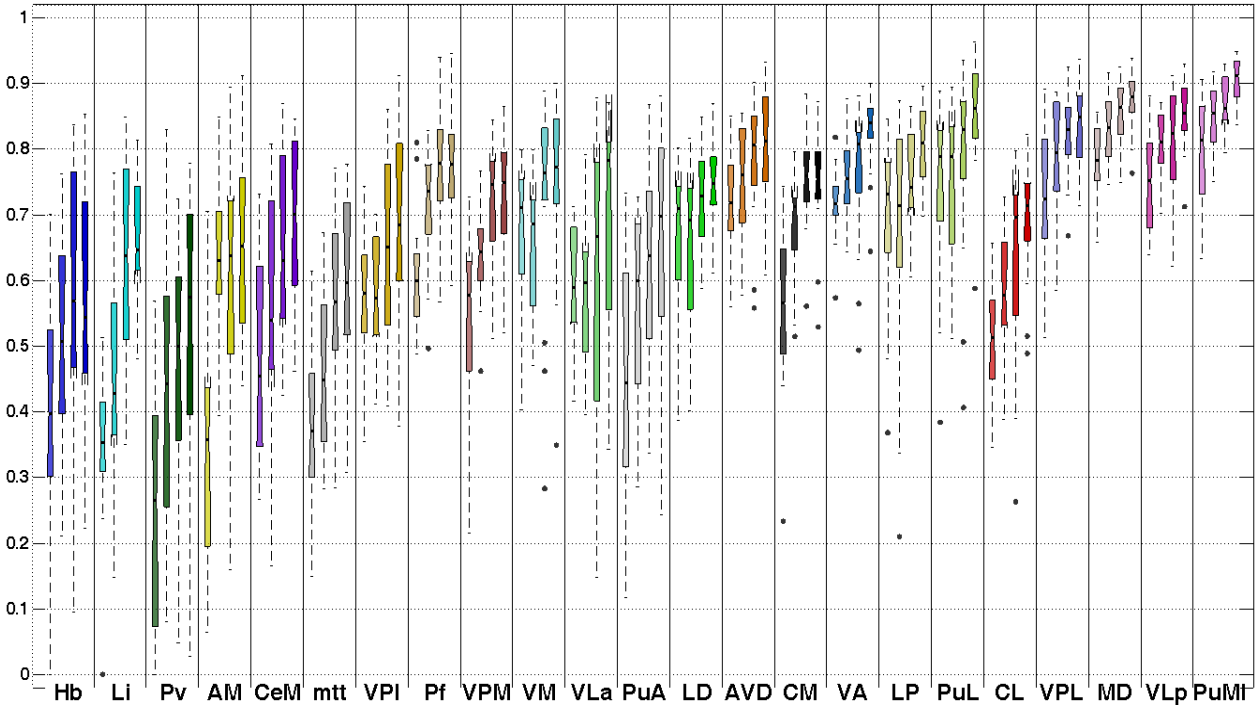

Figure 6.11. Boxplot of dice coefficients for the sixteen test cases in the leave-one-out experiment. Thalamic nuclei are ordered ascendingly according to their mean volumes of manual delineations. For each nucleus, results obtained with SA, MA, NHSSM, and HSSM approaches are shown from left to right respectively.

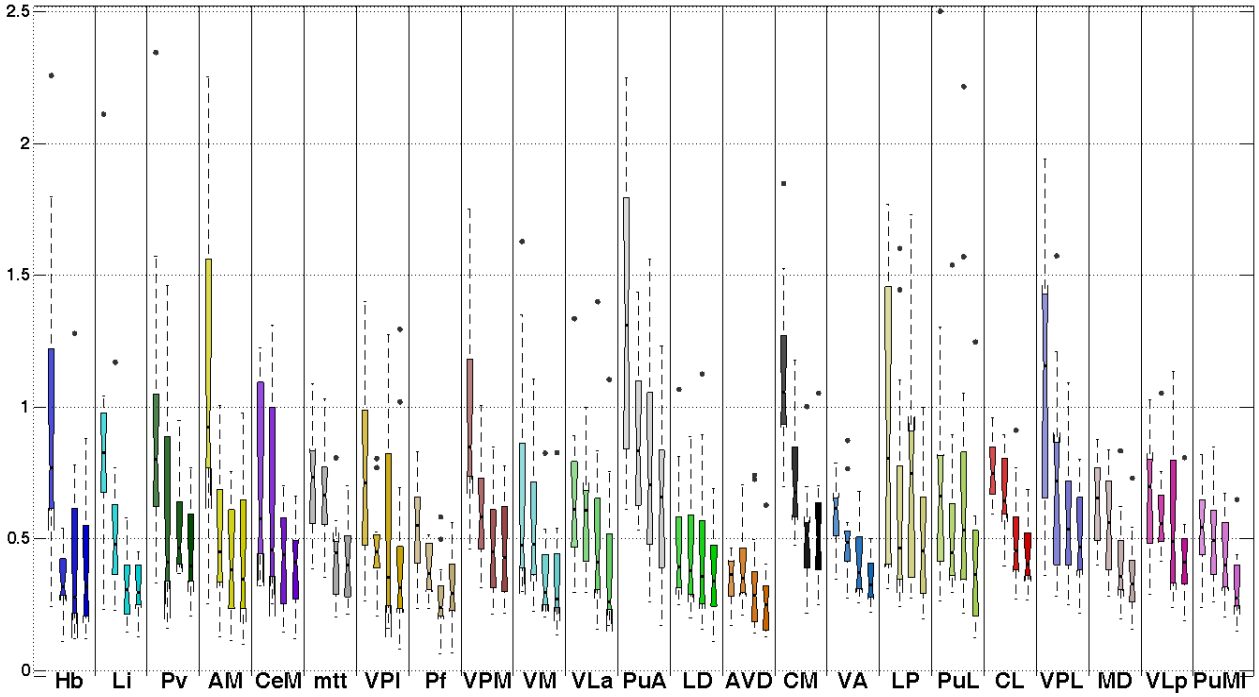

Figure 6.12. Boxplot of mean surface errors (in millimeter) for the sixteen test cases in the leave-one-out experiment. Thalamic nuclei are ordered ascendingly according to their mean volumes of manual delineations. For each nucleus, results obtained with SA, MA, NHSSM, and HSSM approaches are shown from left to right respectively.

Table 6.3. Mean and standard deviation of dice coefficient and mean surface errors for the sixteen test using SA, MA, NHSSM, and HSSM approaches

	Dice Coefficient				Mean Surface Error			
	SA	MA	NHSSM	HSSM	SA	MA	NHSSM	HSSM
Hb	0.39 ±0.20	0.51 ±0.16	0.58 ±0.22	0.57 ±0.18	0.94 ±0.56	0.34 ±0.10	0.43 ±0.31	0.39 ±0.21
Li	0.35 ±0.12	0.45 ±0.18	0.62 ±0.16	0.66 ±0.10	0.87 ±0.39	0.52 ±0.23	0.32 ±0.12	0.31 ±0.09
Pv	0.25 ±0.19	0.43 ±0.20	0.46 ±0.19	0.53 ±0.21	0.92 ±0.50	0.63 ±0.40	0.53 ±0.16	0.45 ±0.15
AM	0.35 ±0.19	0.64 ±0.11	0.61 ±0.19	0.66 ±0.15	1.14 ±0.61	0.50 ±0.23	0.42 ±0.22	0.43 ±0.27
CeM	0.48 ±0.16	0.55 ±0.19	0.66 ±0.15	0.69 ±0.13	0.70 ±0.33	0.64 ±0.38	0.43 ±0.17	0.38 ±0.15
mtt	0.39 ±0.13	0.46 ±0.12	0.57 ±0.14	0.59 ±0.14	0.72 ±0.19	0.67 ±0.18	0.42 ±0.15	0.41 ±0.14
VPI	0.58 ±0.10	0.58 ±0.09	0.65 ±0.14	0.69 ±0.14	0.75 ±0.35	0.47 ±0.15	0.52 ±0.36	0.42 ±0.33
Pf	0.61 ±0.10	0.72 ±0.09	0.78 ±0.09	0.78 ±0.09	0.53 ±0.18	0.38 ±0.10	0.27 ±0.13	0.31 ±0.14
VPM	0.55 ±0.13	0.63 ±0.08	0.71 ±0.10	0.72 ±0.10	0.95 ±0.34	0.59 ±0.18	0.47 ±0.19	0.45 ±0.18
VM	0.67 ±0.12	0.66 ±0.10	0.72 ±0.17	0.75 ±0.15	0.66 ±0.40	0.56 ±0.27	0.36 ±0.16	0.34 ±0.18
VLa	0.59 ±0.09	0.58 ±0.12	0.60 ±0.23	0.68 ±0.19	0.66 ±0.25	0.60 ±0.22	0.49 ±0.31	0.41 ±0.26
PuA	0.45 ±0.20	0.56 ±0.14	0.62 ±0.15	0.67 ±0.17	1.34 ±0.57	0.88 ±0.28	0.79 ±0.40	0.64 ±0.28
LD	0.67 ±0.11	0.65 ±0.13	0.72 ±0.08	0.75 ±0.07	0.49 ±0.24	0.45 ±0.21	0.44 ±0.28	0.37 ±0.18

AVD	0.72 ±0.08	0.75 ±0.08	0.78 ±0.10	0.81 ±0.09	0.34 ±0.09	0.38 ±0.13	0.33 ±0.19	0.27 ±0.13
CM	0.55 ±0.12	0.68 ±0.08	0.75 ±0.08	0.75 ±0.09	1.13 ±0.30	0.74 ±0.21	0.50 ±0.18	0.53 ±0.20
VA	0.71 ±0.06	0.75 ±0.06	0.77 ±0.11	0.82 ±0.06	0.59 ±0.12	0.50 ±0.15	0.41 ±0.12	0.35 ±0.09
LP	0.69 ±0.13	0.68 ±0.19	0.75 ±0.08	0.80 ±0.06	0.92 ±0.54	0.62 ±0.43	0.70 ±0.38	0.49 ±0.24
PuL	0.74 ±0.14	0.74 ±0.13	0.78 ±0.14	0.85 ±0.09	0.76 ±0.54	0.55 ±0.31	0.69 ±0.53	0.40 ±0.27
CL	0.51 ±0.08	0.57 ±0.10	0.63 ±0.15	0.69 ±0.09	0.75 ±0.11	0.68 ±0.15	0.50 ±0.17	0.44 ±0.12
VPL	0.73 ±0.10	0.78 ±0.09	0.82 ±0.07	0.84 ±0.06	1.11 ±0.46	0.72 ±0.36	0.57 ±0.25	0.50 ±0.18
MD	0.78 ±0.06	0.83 ±0.05	0.85 ±0.05	0.87 ±0.05	0.64 ±0.16	0.57 ±0.18	0.41 ±0.17	0.35 ±0.14
VLp	0.75 ±0.07	0.81 ±0.05	0.80 ±0.09	0.85 ±0.06	0.66 ±0.22	0.60 ±0.16	0.56 ±0.28	0.41± 0.15
PuMI	0.80 ±0.08	0.85 ±0.05	0.87 ±0.04	0.90 ±0.03	0.55 ±0.16	0.51 ±0.18	0.43 ±0.15	0.31 ±0.12

Results for thalamic nuclei substructures are shown in ascending order according to their mean volumes of manual delineations.

6.4. Conclusions

In this chapter, we have described a protocol for generating a human atlas of the thalamus using ultra-high field imaging. Superior image resolution and the wide range of contrast elicited from various 7T sequences *in vivo* dramatically improve anatomic delineation of thalamic nuclei and allow the identification of 23 internal substructures in nine healthy subjects on each side of the brain. We have also proposed single-atlas-based, multi-atlas-based, and statistical shape-based methods to use such atlases for thalamic nuclei segmentation in standard 3T T1-w images. From the results of a leave-one-out experiment, we observe that the multi-atlas approach statistically improves upon the single-atlas approach with higher dice values and lower MSE for most nuclei structures, while statistical shape-based method consistently yields the highest segmentation accuracy. Segmentations obtained with SSM agree strongly with the manual delineations and preserve realistic shapes, indicating the potential value of such method to segment the internal nuclei. This opens the door to new solutions for DBS planning and neuroanatomical studies.

Despite promising results, there are several unresolved problems that need to be discussed. In this study, we rely on the set of 7T sequences as a combination to manually

delineate the internal nuclei. None of the images alone is sufficient to localize the entire external boundary of each nucleus. It is desirable to investigate the contribution of each sequence to the visualization of the thalamic structures. Meanwhile, to analyze all the sequences, one needs to examine each of them by constantly flipping from one to another. In the future, a synthetic image that combines the contrast of all the 7T sequences might ease the manual segmentation task.

Due to the lack of an absolute subject-specific gold standard, it is difficult to assess the accuracy of the manual outlines and the subsequent semi-automatic segmentation results. Several directions could be pursued to partially address this issue. Firstly, the reliability of the manual delineations should be quantified by repeating the same segmentation process so as to measure intra- and inter-rater variability of each structure. Secondly, postmortem brains could be scanned using the proposed 7T sequences. Image-derived manual segmentations could then be compared with histological verification of nuclear anatomy in these specimens to further validate our atlas generation protocol. Thirdly, a study of the anatomical-physiological correlations from microelectrode recordings collected during the DBS surgery would be an important step towards clinical validation of the semi-automatic approaches for thalamic structures. Lastly, it would be of interest to compare our thalamic segmentations with those obtained from different data sources such as DTI or fMRI. This may require combining certain substructures in our segmentations, as those techniques often result in 14 structures or less.

In this study, the semi-automatic segmentations of intrathalamic structures are obtained either by means of registrations using atlas-based approaches or by the relationship between the thalamus and the internal substructures inferred from statistical shape models. We believe this is the best we can achieve given only 3T T1-w images. The 3T T1-w sequence, with little contrast within the thalamus region, remains the most common pulse sequence and the standard clinical

imaging for brain surgeries. If additional or alternative sequences are allowed, image forces derived from such sequences may be used to improve the segmentation accuracy. For example, Stough *et al.* extract multi-modal features from both MPRAGE and DTI to learn the location of six nuclei structures [35].

In the statistical shape-based segmentations, we find that constructing multi-level shape models yields better results that using a single model. Although both methods rely on the external thalamus surface to infer the internal structures, the multi-level models may provide better shape representation by delegating variations within the nuclei complex to the lower levels and capturing limited shape variations at each level. This becomes useful when there are not sufficient training shapes, as is the case in our study. To further compensate for the limited training shapes, we build each shape model by pooling from both sides of the brain. However, anatomical variations of the thalamic structure that are specific to each hemisphere may exist. It may be favorable to acquire more training subjects and build customized models for each hemisphere.

A critical component in shape-based approach is the homologous point identification between training shapes. This is a challenging task, especially for 3D structures. Currently, this is done for each structure separately using various registrations and occasional manual adjustment to align the reference shapes with the target shapes. Optimizing point correspondence structure by structure may not be optimal, as adjacent points that belong to two structures in the reference shapes may end up being further apart in the target shapes. Alternative approaches that optimize point correspondence jointly for all structures should be explored.

Furthermore, this is a study of healthy volunteers. Atlases generated from those subjects may have limited capability in expressing potential pathology in thalamic regions. It would be

beneficial to study shape variations in thalamic nuclei structures between normal control and the actual patients. Moreover, successful clinical application of our approach to DBS patients, which is one of our ultimate goals, may require reconstruction of thalamus atlases from the patient population and adjustment in the segmentation approaches to account for pathological changes. We plan to recruit a number of DBS patients to regenerate thalamus atlases for this purpose in the future.

References

[1] G. M. Halliday, "Thalamic changes in Parkinson's disease," *Parkinsonism Relat Disord*, vol. 15, pp. S152-S155, 2009.

[2] A. Cifelli, M. Arridge, P. Jezzard, M. M. Esiri, J. Palace, and P. M. Matthews, "Thalamic neurodegeneration in multiple sclerosis," *Ann Neurol*, vol. 52, no. 5, pp. 650-653, 2002.

[3] A. V. Apkarian, Y. Sosa, S. Sonty, R. M. Levy, R. N. Harden, T. B. Parrish, and D. R. Gitelman, "Chronic back pain is associated with decreased prefrontal and thalamic gray matter density," *J Neurosci*, vol. 24, no. 46, pp. 10410-10415, 2004.

[4] W. Koller, R. Pahwa, K. Busenbark, J. Hubble, S. Wilkinson, A. Lang, and T. Malapira, "High-frequency unilateral thalamic stimulation in the treatment of essential and parkinsonian tremor," *Ann Neurol*, vol. 42, no. 3, pp. 292-299, 1997.

[5] A. Morel, *Stereotactic atlas of the human thalamus and basal ganglia*, CRC Press, 2013.

[6] K. Niemann, V. R. Mennicken, D. Jeanmonod, and A. Morel, "The Morel stereotactic atlas of the human thalamus: atlas-to-MR registration of internally consistent canonical model," *Neuroimage*, vol. 12, no. 6, pp. 601-616, 2000.

[7] K. A. Ganser, H. Dickhaus, R. Metzner, and C. R. Wirtz, "A deformable digital brain atlas system according to Talairach and Tournoux," *Med Image Anal*, vol. 8, no. 1, pp. 3-22, 2004.

[8] A. Krauth, R. Blanc, A. Poveda, D. Jeanmonod, A. Morel, and G. Székely, "A mean three-dimensional atlas of the human thalamus: generation from multiple histological data," *Neuroimage*, vol. 49, no. 3, pp. 2053-2062, 2010.

[9] M. R. Wiegell, D. S. Tuch, H. B. Larsson, and V. J. Wedeen, "Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging," *Neuroimage*, vol. 19, no. 2, pp. 391–401, 2003.

[10] U. Ziyan, D. Tuch, and C. F. Westin, "Segmentation of thalamic nuclei from DTI using spectral clustering," in *Proc. MICCAI*, 2006, pp. 807-814.

[11] Y. Duan, X. Li, and Y. Xi, "Thalamus segmentation from diffusion tensor magnetic resonance imaging," *Int J Biomed Imaging*, no. 7, pp. 1-5, 2007.

[12] T. E. Behrens, *et al.*, "Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging," *Nature neurosci*, vol. 6, no. 7, pp. 750–757, 2003.

[13] H. Johansen-Berg, T. E. Behrens, E. Sillery, O. Ciccarelli, A. J. Thompson, S. M. Smith, and P. M. Matthews, "Functional–anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus," *Cereb Cortex*, vol. 15, no. 1, pp. 31-39, 2005.

[14] D. Zhang, A. Z. Snyder, J. S. Shimony, M. D. Fox, and M. E. Raichle, "Noninvasive functional and structural connectivity mapping of the human thalamocortical system," *Cereb Cortex*, vol. 20, no. 5, pp. 1187–1194, 2010.

[15] C. R. Traynor, G. J. Barker, W. R. Crum, S. C. Williams, and M. P. Richardson, "Segmentation of the thalamus in MRI based on T1 and T2," *Neuroimage*, vol. 56, no. 3, pp. 939-950, 2011.

[16] S. C. Deoni, M. J. Josseau, B. K. Rutt, and T. M. Peters, "Visualization of thalamic nuclei on high resolution, multi-averaged T1 and T2 maps acquired at 1.5 T," *Hum Brain Mapp*, vol. 25, no. 3, pp. 353-359, 2005.

[17] A. Sudhyadhom, I. U. Haq, K. D. Foote, M. S. Okun, and F. J. Bova, "A high resolution and high contrast MRI for differentiation of subcortical structures for DBS targeting: the Fast Gray Matter Acquisition T1 Inversion Recovery (FGATIR)," *Neuroimage*, vol. 47, pp. T44-T52, 2009.

[18] F. Vassal, J. Coste, P. Derost, V. Mendes, J. Gabrillargues, C. Nuti, and J. J. Lemaire, "Direct stereotactic targeting of the ventrointermediate nucleus of the thalamus based on anatomic 1.5-T MRI mapping with a white matter attenuated inversion recovery (WAIR) sequence," *Brain Stimul*, vol. 5, no. 4, pp. 625-633, 2012.

[19] T. Tourdias, M. Saranathan, I. R. Levesque, J. Su, and B. K. Rutt, "Visualization of intrathalamic nuclei with optimized white-matter-nulled MPRAGE at 7T," *Neuroimage*, vol. 84, no. 1, pp. 534-545, 2014.

[20] A. Abosch, E. Yacoub, K. Ugurbil, and N. Harel, "An assessment of current brain targets for deep brain stimulation surgery with susceptibility-weighted imaging at 7 tesla," *Neurosurgery*, vol. 67, pp. 6, pp. 1745-1756, 2010.

[21] A. T. Newton, B. M. Dawant, and P.-F. D'Haese, "Visualizing intrathalamic structures with combined use of MPRAGE and SWI at 7T," in *Proc. ISMRM*, 2014, pp. 122.

[22] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput Vis Image Und*, vol. 61, no. 1, pp. 38-59, 1995.

[23] R. Blanc, C. Seiler, G. Sz ékely, L. P. Nolte, and M. Reyes, "Statistical model based shape prediction from a combination of direct observations and various surrogates: application to orthopaedic research," *Med Image Anal*, vol. 16, no. 6, pp. 1156-1166, 2012.

[24] N. Baka, C. Metz, M. Schaap, B. Lelieveldt, W. Niessen, and M. de Bruijne, "Comparison of shape regression methods under landmark position uncertainty," in *Proc. MICCAI*, 2011, pp. 434-441.

[25] J. Kim, Y. Duchin, G. Sapiro, J. Vitek, and N. Harel, "Clinical deep brain stimulation region prediction using regression forests from high-field MRI," in *Proc. ICIP*, 2015, pp. 2480-2484.

[26] A. R. Khan, L. Wang, and M. F. Beg, "FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping," *Neuroimage*, vol. 41, no. 3, pp. 735-746, 2008.

[27] F. Maes, F. A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans Med Imaging*, vol. 16, no. 2, pp. 187-198, 1997.

[28] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int J Comput Vis*, vol. 24, no. 2, pp. 137-154, 1997.

[29] G. K. Rohde, A. Aldroubi, and B. M. Dawant, "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Trans Med Imaging*, vol. 22, no. 11, pp. 1470-1479, 2003.

[30] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int J Comput Vis*, vol. 1, no. 4, pp. 321-331, 1988.

[31] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans Pattern Anal Mach Intell*, vol. 11, no. 6, pp. 567-585, 1989.

[32] K. Rohr, H. S.Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Trans Med Imag*, vol. 20, no. 6, pp. 526-534, 2001.

[33] J. A. Sethian, *Level Set Methods and Fast Marching Methods*, 2nd ed. Cambridge, MA: Cambridge Univ. Press, 1999.

[34] B. F. Green, "The orthogonal approximation of an oblique structure in factor analysis," *Psychometrika*, vol. 17, no. 4, pp. 429-440, 1952.

[35] J. V. Stough, J. Glaister, C. Ye, S. H. Ying, J. L. Prince, and A. Carass, "Automatic method for thalamus parcellation using multi-modal feature classification," in *Proc. MICCAI*, 2014, pp. 169-176.

CHAPTER VII

SUMMARY AND FUTURE WORK

This dissertation introduces several innovative algorithms to improve surgical guidance for preoperative planning in DBS procedures. Specifically, these methods are developed and evaluated to address two tasks, namely, selecting an ideal trajectory for electrode implantation as described in Chapter II and localizing an accurate position for target stimulation as described in Chapter III, IV, V, and VI.

In Chapter II, we propose a trajectory planning system and a validation approach for a thorough clinical evaluation. This system finds an optimal path for inserting the electrode by simultaneously analyzing a large number of candidate trajectories against many surgical constraints, ranking them according to a weighted linear combination of those constraints, and choosing the trajectory with lowest cost as the optimal solution. Weights for those constraints are manually initialized and heuristically refined to capture the individual neurosurgeon preference in selecting trajectories. Through a series of retrospective experiments conducted internally, we observe that one universal set of weights is preferable to those that are customized for different neurosurgeons. Preceded with this set of weights, results of those experiments show that automatic trajectories generated by our system outperform the manual selections both qualitatively and quantitatively. We further evaluate the automatic trajectories in a large-scale pseudo-prospective study involving four neurosurgeons at three institutions. Out of 60 cases, automatic trajectories are accepted 95% of the time by a majority of neurosurgeons and 76.67%

unanimously. This suggests the clinical applicability of the proposed trajectory planning system across multiple sites and surgeons.

In Chapter III, we develop a learning-based approach to localize the AC, PC, and MSP so that a spatial coordinate system can be established to indirectly determine the target position. We describe a voxel-based regression learning framework that relates the appearance of a point to the probability of this point being a landmark or in the plane. Regression forests models are built at three resolution levels and applied hierarchically to localize the AC, PC, and MSP. The final position of the landmarks and the plane are determined by the mean shift algorithm and least square fitting respectively in a weighted fashion. We design three experimental paradigms to evaluate the performance of our framework, which include a large-scale leave-one-out experiment, a series of robustness tests, and parameter sensitivity studies. Results of the first study shows that our method leads to an overall error of 0.55 ± 0.30 mm for AC, 0.56 ± 0.28 mm for PC, $1.08^{\circ} \pm 0.66^{\circ}$ in the plane's normal direction and 1.22 ± 0.73 voxels in average displacement for MSP. This compares favorably to most state-of-the-art methods including six registration-based methods and a publicly available toolkit. Our method also performs well under varying degrees of spherical lesions, tissue deformations, imaging noise and rotations. It takes about 25 seconds on a standard PC to detect all objects at once. These experiments demonstrate the accuracy, robustness, and efficiency of our method for AC, PC, and MSP localization, and thus for indirect targeting.

For statistical atlases-based targeting, nonrigid registration is the key to form an accurate mapping between patient volumes and a reference template. In Chapter IV, we investigate the effect of nonrigid registration algorithm on the creation of statistical atlases. A total of eighteen statistical atlases are created by registering 100 patients all targeting the STN to each of the three

reference templates using six well-known nonrigid registration algorithms. Through qualitative comparisons, probabilistic maps generated from those atlases all cover a similar region of the STN. Among these algorithms, the atlases share similar centroids and spreads with sub-voxel differences. This indicates that the choice of registration algorithm does not significantly impact the quality of statistical atlases in the region of the STN.

In Chapter V, we present an alternative targeting approach by directly learning the target position from past patients. This is achieved by learning a multi-variant regression forests model that associates the description of each point with a displacement vector of this point to the actual target. Each point is described by its spatial coordinate in the AC-PC space calculated using the technique proposed in Chapter III as well as intensity contextual features extracted from T1-w and T2-w sequences. Target position is finalized by aggregating predictions made by the model of all voxels in a weighted fashion. Through a five-fold cross validation with 100 subjects, we demonstrate the role of spatial features in regularizing the predictions and the advantage of using both modalities over either one alone in the targeting accuracy. With 2.63 ± 1.37 mm targeting errors, we show that our approach serves as an accurate and robust option for targeting compared to statistical atlas-based methods and AC-PC-based methods.

In Chapter VI, we design a protocol for building high-field thalamic atlases and proposed methods for using such atlases to segment thalamic substructures on clinical data. These atlases involve a set of 7T MR images with complementary contrast acquired for each subject and manually delineated thalamic nuclei based on the Morel atlas convention. To segment these structures in clinical 3T images, we apply single-atlas-based and multi-atlas-based approaches that rely on registration between the atlas and the subject. We also propose a statistical shape-based approach that builds joint shape models of the thalamus and the internal nuclei and

hierarchically segments the nuclei by fitting the thalamus to those models. Leave-one-out validation performed on the sixteen test cases shows that the three methods yield reasonable results. Segmentations obtained with shape-based approach agree the strongest with manual delineations while preserving realistic shapes; the dice coefficients range from 0.53 to 0.90 and mean surface errors from 0.34 mm to 0.88 mm for each structure averaged across test cases. This suggests the feasibility of the shape-based approach using those atlases for thalamic nuclei segmentation, and thereby for DBS targeting with regard to thalamic structures.

Even though we have made substantial progress towards developing computer-aided automatic approaches for DBS surgical planning, work remains. With regards to trajectory planning, it is of note that the weighting parameters used to score trajectories are set heuristically in the current platform. Due to the lack of ground truth, it is challenging to determine the optimal choice of weights for different surgical constrains. Essert *et al.* recently proposed two approaches to retrieve the weights that best matches the neurosurgeon's manual planning, one based on a stochastic sampling of the parameters and the other on an exhaustive search [1]. For clinical use, a system that allows interactive adjustment of the weights may be more helpful. The trajectory is commonly represented as a straight line to model the single track electrode. Under other clinical settings, it might be beneficial to adopt a cylinder-shape representation as done by Bériault *et al.* [2] and even consider the multi-track electrode device.

One underlying assumption of the trajectory planning system is a perfectly chosen target point. However, modification of the target position happens, which may invalidate the pre-computed trajectory. To avoid another insertion, uncertainty in the target prediction, especially the shape of the uncertainty region, should be taken into account. Clinical implementation that

pre-computes a series of solutions or constraining the search space may also be required to produce new solutions in clinically acceptable time.

Several research directions can be considered to further improve the targeting accuracy. As suggested by Guo *et al.*, targeting with comprehensive anatomical and functional information results in estimates that are much closer to the optimal surgical target [3]. We have proposed several targeting approaches with different working mechanisms that exploit various data sources such as the segmentation or localization of anatomical structures or actual surgical target collection. Except for the one in Chapter VI, the proposed techniques are generic regardless of target structures and have been extensively validated for STN targeting. It would be helpful to evaluate their performance on thalamic nuclei targeting and, furthermore, to combine these algorithms for a more accurate and robust estimate.

Advances in imaging inexorably have led, and will further lead, to better preoperative delineation of the targets [4]. From 1.5T to 3T as the current clinical standard to the recently introduced 7T imaging, tremendous progress has been and is actively being made in defining targets on these scanners. A major concern with 7T MR imaging is the greater tissue distortion and susceptibility to artifacts at high field strength. Once accepted for clinical use, we believe 7T datasets will play a critical role in DBS targeting.

While imaging technologies continue to evolve, the gold standard for clinically verifying the target position remains intraoperative MER. Furthermore, there exists a potential discrepancy between decisions made by imaging and those by neurophysiology, which may not be resolved in the preoperative stage [5]. Intraoperative guidance towards adjusting and finalizing the electrode placement can be extremely valuable. Relevant past efforts concentrate on developing advanced signal processing techniques to better characterize the functional borders of target

structures [6-7]. These are mostly static methods that use only MER data. Instead, a reinforcement learning paradigm that exploits preoperative information and the streaming intraoperative MER data may be possible to adapt the electrode insertion automatically. Specifically, as the surgery progresses, instructions can be made to inform the surgeons about the next stimulation point in real time based on the image data as well as current recordings while accounting for the allowable region the electrode can explore. This may serve as a great addition to the current surgical guidance system.

Another possible source of inaccuracy in the preoperative targeting is the intraoperative brain shift. Brain shift, mainly caused by CSF loss, can have a considerable impact on the final electrode position [8]. Quantification of the brain shift is difficult due to the absence of intraoperative imaging. Pallavaram *et al.* proposed to detect brain shift by correlating low-shift electrophysiological maps with intraoperative observations [9]. Future work that corrects for brain shift is necessary to enhance the clinical utility of preoperative planning.

We are currently underway to integrate the methods proposed in this work into a clinical processing pipeline at Vanderbilt. Once completed, prospective evaluation studies will be conducted. Although the solutions that have been presented herein may not be the final ones, we hope that these works have made valuable contributions towards solving the problem of DBS preoperative planning, and possibly serving as inspirations for other image-guided therapy issues.

References

[1] C. Essert, S. Fernandez-Vidal, A. Capobianco, C. Haegelen, C. Karachi, E. Bardinet, M. Marchal, and P. Jannin, "Statistical study of parameters for deep brain stimulation automatic preoperative planning of electrodes trajectories," *Int J Comput Assist Radiol Surg*, vol. 10, no. 12, pp. 1973-1983, 2015.

[2] S. Bériault, F. A. Subaie, D. L. Collins, A. F. Sadikot, and G. B. Pike, "A multi-modal approach to computer-assisted deep brain stimulation trajectory planning," *Int J Comput Assist Radiol Surg*, vol. 7, no. 5, pp. 687–704, 2012.

[3] T. Guo, A. G. Parrent, and T. M. Peters, "Surgical targeting accuracy analysis of six methods for subthalamic nucleus deep brain stimulation," *Compute Aided Surg*, vol. 12, no. 6, pp. 325-334, 2007.

[4] R. E. Gross and M. E. McDougal, "Technological advances in the surgical treatment of movement disorders," *Curr Neurol Neurosci*, vol. 13, no. 8, pp. 1-12, 2013.

[5] M. Bot, L. Bour, R. de Bie, M. F. Contarino, R. Schuurman, and P. Munckhof van den, "Can we rely on susceptibility-weighted imaging for subthalamic nucleus identification in deep brain stimulation surgery?" *Neurosurgery*, vol. 78, no. 3, pp. 353-360, 2016.

[6] R. R. Shamir, A. Zaidel, L. Joskowicz, H. Bergman, and Z. Israel, "Microelectrode recording duration and spatial density constraints for automatic targeting of the subthalamic nucleus," *Stereotactic Funct Neurosurg*, vol. 90, no. 5, pp. 325-334, 2012.

[7] S. Wong, G. H. Baltuch, J. L. Jaggi, and S. F. Danish, "Functional localization and visualization of the subthalamic nucleus from microelectrode recordings acquired during DBS surgery with unsupervised machine learning," *J Neural Eng*, vol. 6, no. 2, pp. 026006, 2009.

[8] Y. Miyagi, F. Shima, and T. Sasaki, “Brain shift: an error factor during implantation of deep brain stimulation electrodes,” *J Neurosurg*, vol. 107, no. 5, pp. 989-997, 2007.

[9] S. Pallavaram, P.-F. D’Haese, M. S. Remple, J. S. Neimat, C. Kao, R. Li, P. E. Konrad, and B. M. Dawant, “Detecting brain shift during deep brain stimulation surgery using intraoperative data and functional atlases: a preliminary study,” in *Proc. ISBI*, 2009, pp. 362-365.