SPINAL CORD SEGMENTATION AND DISABILITY PREDICTION IN MULTIPLE SCLEROSIS USING NOVEL OPTIMIZATION AND MACHINE LEARNING METHODS

by

Jeremy Kawahara

B.Sc., Vancouver Island University, 2011

A Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Summer 2013

APPROVAL

Name: Jeremy Kawahara

Degree: Master of Science

Title of Thesis: Spinal cord segmentation and disability prediction in multiple sclerosis using novel optimization and machine learning methods

Examining Committee: Dr. Andrei Bulatov
Chair

Dr. Ghassan Hamarneh, Senior Supervisor,
Associate Professor

Dr. Roger Tam, Supervisor,
Assistant Professor, Radiology,
The University of British Columbia

Dr. Chris McIntosh, Supervisor,
Research Associate,
Princess Margaret Cancer Centre

Dr. Mehdi Moradi, External Examiner,
Assistant Professor,
Electrical and Computer Engineering,
The University of British Columbia

Date Approved: August 9, 2013

Partial Copyright Licence

Abstract

Multiple sclerosis studies show a correlation between spinal cord atrophy and physical disability, indicating the potential use of atrophy as a biomarker to monitor disease progression. To quantify spinal cord atrophy, clinicians need to accurately measure the cord and determine which cord properties consistently capture tissue loss. We address these needs by making three contributions: (i) a novel algorithm to segment the spinal cord by finding the globally optimal minimal path in six dimensions; (ii) a machine learning spinal cord segmentation approach where we introduce the concept of global geometric features into decision forests to address the first algorithm's limitations; and (iii) novel morphological and appearance features extracted from magnetic resonance images (MRI) and corresponding spinal cord segmentations that are combined to predict the physical disability of patients with multiple sclerosis. Our results demonstrate improvements over state-of-the-art spinal cord segmentation methods and improved prediction of clinical disability from MRI data.

Keywords: Spinal cord segmentation; multiple sclerosis; minimal path optimization; machine learning; feature extraction; regression

To Emma

Acknowledgements

There are many people whose support I am very grateful for.

Thank you Dr. Ghassan Hamarneh, my senior supervisor whose work ethic and passion for medical image analysis inspires me; Dr. Roger Tam, my supervisor whose kind words of encouragement and knowledgeable expertise has taught me so much; and, Dr. Chris McIntosh, my supervisor whose excitement for research and deep understanding of the field has shaped how I now think.

Thank you to all my labmates, for our many hours of in-depth discussions. Specifically, thank you Tayebeh Lotfi for all the time you took sharing your knowledge with me.

Thank you Biogen Idec Canada, for funding a large portion of this work and giving me the opportunity to focus solely on this research.

I am grateful to, my father Glen, for patiently and calmly listening to me in all my moments of weaknesses; my mother Jacoba, for always being excited for everything that I do; and my sisters, Liz, Steph, Maria and Anna, for being a source of my motivation.

And thank you Emma Leach, for constantly reminding me that we are not robots, yet.

Contents

List of Tables

List of Figures

Chapter 1

Background

1.1 Multiple Sclerosis

Multiple sclerosis (MS), named after the multiple scars or lesions it causes, is a disease that inflames and damages a patient's central nervous system [11, 14, 54], which is comprised of the brain and spinal cord. This disease is one of the most common causes of neurological disability in young adults [14, 54]. In Canada, the overall prevalence of MS is reported as 240 per 100,000 people [7].

The causes of MS are unknown but it is thought to be influenced by a combination of genetic and environmental factors [7]. MS, a neurological disease associated with autoimmune dysfunction, has been hypothesised to be an autoimmune disease, and more recently a disorder with the cells that create the myelin sheath (i.e. oligodendrogliopathy) [54]. There is no known cure for MS and the current disease modifying therapies used for treatment have only demonstrated modest effects [54].

Symptoms of MS commonly include disability to the ambulatory, upper body and cognitive functions [22]. While the disease can evolve differently for each patient, there are common patterns in the progression. After the onset of MS, a patient will either have relapse-remitting MS marked by periods of neurological recovery during the remittance period, or a patient will have a primary progressive form of MS where there are no recovery periods. For patients with relapse-remitting MS, the disease may progress to secondary progressive MS where the recovery periods lessen and the neurological functions continue to decrease [54], or a patient will stabilize to a benign form of MS [57].

As MS progresses a clinician will administer a series of tests to measure the patient's

functional capabilities. The results of these tests are used to form a numerical score giving a quantifiable indication of the patient's disability level. Common tests are the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) score. The EDSS score has a heavy focus on the ambulatory disability of a patient while the MSFC score equally considers ambulatory, upper body and cognitive functions [22].

Studies have shown that the spinal cord, a region of the central nervous system imperative to physical function, is often damaged in patients with MS and thus is thought to play a significant role in the displayed physical symptoms of the disease [21]. For in vivo analysis of the pathology of MS, magnetic resonance imaging has proven to be a valuable tool [3].

1.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a way to visualize the internal structural information of a patient. To accomplished this the body is placed within a magnetic field which causes the axes of hydrogen protons to align. A specific radio frequency is applied to the magnetic field causing the hydrogen nuclei to resonate. When this frequency is removed, the protons align to their resting positions (a process known as relaxation) and emit a signal. The intensity of the signal is captured and converted to a grey scale image. Different tissues can be identified by their different relaxation times which produce different signals [9, 16, 27].

This relaxation process is known as T_1 and T_2 relaxation. T_1 relaxation, or the spin-lattice relaxation time, is defined as the time it takes for the longitudinal magnetization to recover its initial value. T_2 relaxation, or the spin-spin relaxation time, is defined as the time it takes for the transverse magnetization component to return to equilibrium [27,28]. These relaxation methods result in tissue being visualized differently. To optimize the contrast between the cord and the surrounding cerebrospinal fluid, and the reproducibility of the results, it has been suggested to use 3D T_1-weighted acquisitions [49]. In this work, we examine T_1-weighted scans.

The magnetic field used to align the protons in MR images is also an important consideration. A stronger field can align more protons to increase the signal-to-noise ratio, resolution or speed of the scan. This field strength is measured in teslas (T). In clinical practice, the most common magnetic field strength is 1.5T with 3.0T becoming more common [58]. In this work we use scans acquired with magnetic field strengths of 1.5T and 3.0T.

MRIs are not known to expose patients to biological harm [27], making this a desirable

procedure for patients who require frequent imaging. Furthermore, in studies of MS, MRI has been shown to be a sensitive modality capable of detecting changes in the integrity of brain and spinal cord tissue [3, 42].

There are many extensions to conventional MRI including: diffusion tensor MRI [6], which measures the motion of the water molecules to determine the orientation of axons and the direction of white matter bundles; and functional MRI, which measures the blood-oxygenation-level-dependent contrast mechanism [3]. While there are many promising extensions to MRI, in this work we restrict our focus to conventional structural MR imaging of the spinal cord.

1.3 Computer-aided Spinal Cord Segmentation

Advances in imaging technologies have produced more images at a higher resolution, yielding vast amounts of data that are now beginning to overwhelm radiologists' capacities to analyze them [70]. To assist the radiologists in their analyses, it is critical to develop computed-assisted methods. To aid radiologists in measuring the spinal cord, we develop computer algorithms capable of labelling voxels as spinal cord or background in a 3D MRI. This process of labelling an image is known as *segmentation*.

Segmenting the spinal cord is a particularly challenging problem as the signal changes over the length of the cord, images are often noisy due to patient movement, the boundary of the cord is not well defined in regions where it contacts the wall of the spinal canal, and the cord's shape can change as the disease progresses. Additionally, the changes in the spinal cord from MS are generally subtle. Over the course of a year, the observed change in the area of the spinal cord of an MS patient is approximately 1% - 5% [41, 61, 62], a challenge even for the human eye to detect by only visual inspection. The typical voxel resolution of 1 mm^3 gives ambiguity in the exact location of the boundary of the spinal cord resulting in the boundary voxels being partially labelled as both spinal cord and background, i.e. partial volume (Fig. 1.1). Given the small size of the cord, these partial volumes pixels compose about 25% of the cord area making them important to consider [66]. Segmentation with enough accuracy to detect these minute changes remains a challenge. These challenges combined with the need to minimize the radiologists' and technicians' workload must be taken into account when developing spinal cord segmentation algorithms.

A number of computer-aided methods for spinal cord segmentation have been developed

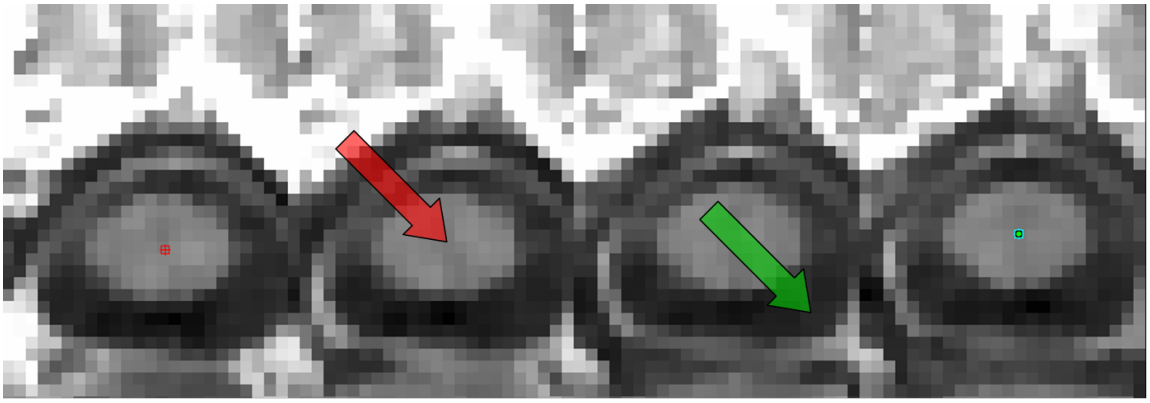
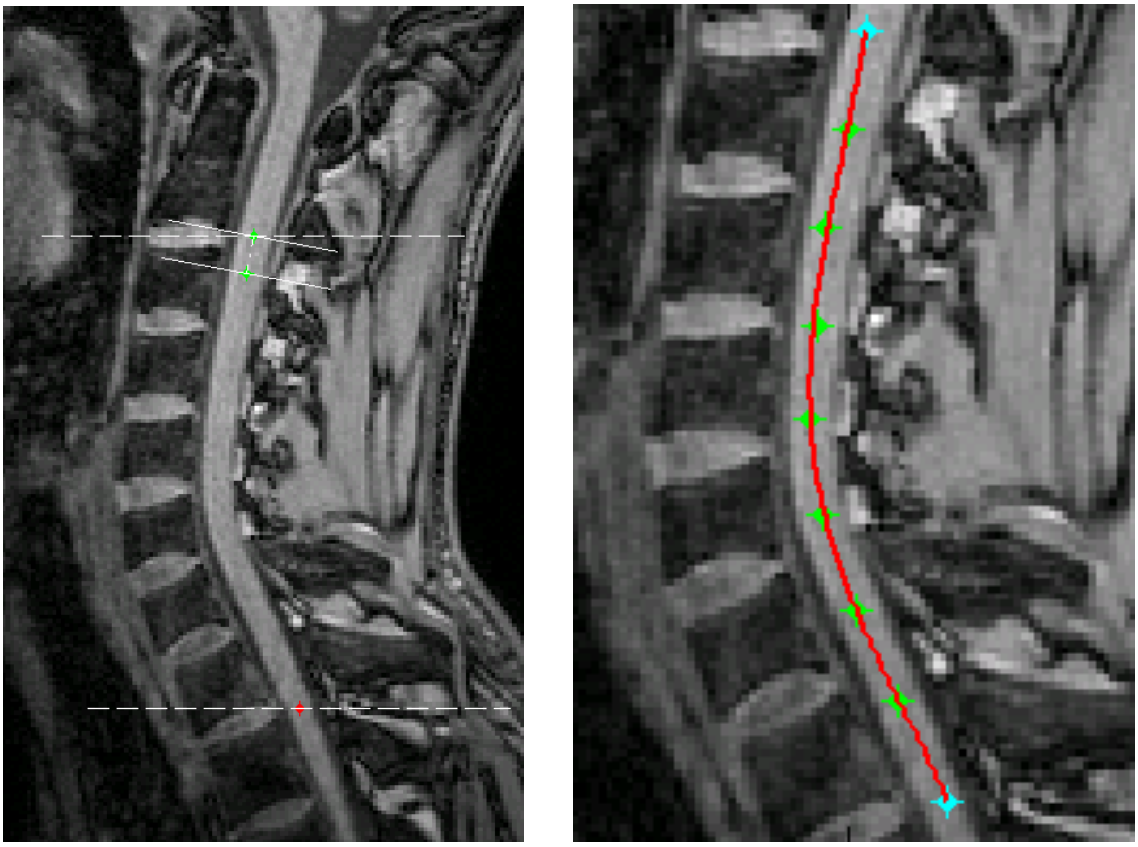


Figure 1.1: Screenshots from the Jim software [1] showing sequential axial slices of the spinal cord with user-marked center points (*small red and green box in the center of the first and last slices*). Note the partial volumes at the border between the spinal cord (*second slice - red arrow*) and the spinal canal (*third slice - green arrow*).

with varying levels of required human interaction. The most basic method would be to manually outline the spinal cord on each 2D slice. Besides being time consuming, this approach has poor scan-rescan reproducibility with a coefficient of variation (CoV, standard deviation divided by the mean) of 6% [42] as applied to images of normal subjects. Highly interactive methods, such as the one proposed by Tench et al. [66], while thought to be sufficiently accurate, still require significant use of manual intervention. The user places seeds points and identifies the appropriate cord boundaries in ambiguous regions. This does produce a flexible spinal cord segmentation, but demands significant amounts of human time (similar methods taking approximately 25 minutes for a single scan [48]). Other methods decrease the amount of human involvement, from requiring seed points on only a few representative slices [29] (Fig. 1.1) to adjusting an automatically computed estimate of the medial path of the spinal cord [48] (Fig. 1.2). However, as the necessary human interaction decreases, often we find that the accuracy also decreases. Parameter tuning or other manual adjustments are generally needed to makes these methods work in a reasonable fashion. A further consideration is that when subjective human input is required, there is also the potential to suffer from intra and inter operator variability. Depending on the method used, intra-operator variability has been reported to range in CoV from 0.44% to 2.15% and inter-operator variability from 1.07% to 7.95% [29]. These factors make the problem of spinal cord segmentation still a very open one. We review the existing spinal cord segmentation algorithms in more detail in Section 2.1 to better contrast existing methods with our approaches.



(a) User entered seed points

(b) Adjustable live-wire path

Figure 1.2: Screenshots from the method proposed by McIntosh et al. [48] demonstrating (a) the user-entered seed points and (b) the live-wire [5] medial path.

The ideal spinal cord segmentation algorithm maximizes the accuracy, maximizes reproducibility, minimizes the user input, and runs in a reasonable amount of time. In this work, we push the boundary of state-of-the-art spinal cord segmentation methods towards one that balances these requirements, with the goal of producing robust segmentations whose extracted measurements will allow us to monitor the spinal cord's subtle changes.

1.4 Models to Predict Physical Disability of MS Patients

Existing segmentation methods, although laborious and possibly suffering in accuracy and precision from the previously mentioned problems, have been used in several studies to report a correlation between spinal cord atrophy and a patient's physical disability [8,29,41, 42,55,57,76]. Spinal cord atrophy is defined as the loss of cord tissue and in many studies is measured by the volume of the segmented spinal cord or by the averaged cross-sectional area (CSA). This assumes the reduction of spinal cord volume (or area) is a good indicator of the true tissue loss. However, the correlations between the measurements taken from MRI and clinical disability outcomes are limited [21]. This limited correlation may be partly due to segmentation error as discussed in Section 1.3, where improved segmentations may improve correlations. But, given the relatively high number (25% [66]) of ambiguous partial volume border voxels relative to the total cord size and the low amount of expected cord change in a year due to MS (1-5% [41,61,62]), even with accurate segmentations the low voxel resolution may prevent us from reaching a useful correlation level between CSA and disablity. This motivates the study of other non-CSA features that are more robust to segmentation error and spinal cord ambiguity. As well, since a single spinal cord feature may not be sufficient to predict a patient's physical disability, it is reasonable to examine methods of combining multiple extracted features from the spinal cord in a regression model.

The importance of spinal cord analysis is further supported since the disease in the spinal cord can progresses independently of the disease in the brain [4]. In the works of Nijeholt et al. [55], they combined brain and spinal cord MRI parameters to see if it improved correlation with multiple sclerois clinical measures. Their results suggested that the brain and the spinal cord both make independent contributions to disability and that the spinal cord symptoms determined the EDSS scores to a large extent. This suggests the potential to model the physical disability of a patient using spinal cord features.

A model that incorporates multiple explanatory variables may provide clinicians with

new insights into the disease and be a useful tool in clinical diagnosis. In addition, novel features that have a high correlation with disability level may prove to be useful biomarkers and used to more effectively monitor the disease.

1.5 Contributions

We hypothesize that 1) more accurate spinal cord segmentations from current MRI data can be produced using high-dimensional path optimization and machine learning methods; and 2) a machine learning model with multiple novel morphological and appearance features will outperform a model using the volume of the cord alone.

To contribute to the testing of these hypotheses, we present the following three works and explain how they are related. In our *first work* we propose a novel spinal cord segmentation method that detects the cord using a hand-crafted energy function with a learned shape regularizer, optimized with a globally optimal high dimensional search. In order to optimize globally we make assumptions to simplify our model; however, we find that our simplifying assumptions made in our hand-crafted energy terms, shape regularization, and optimizer limit the robustness and accuracy of this approach. These limitations motivate the development of our *second work* where we present a machine learning approach to spinal cord segmentation. This machine learning algorithm jointly learns the intensity, regularization and high-level geometric distributions in a non-parametric approach and we demonstrate that this outperforms our original method. Finally our *last work* focuses on the practical use of the segmentations by extracting features from spinal cord segmentations along with the corresponding MRIs to form a regression model capable of predicting the physical disability of a patient. We now give a more detailed overview of each our approaches below.

1.5.1 Globally Optimal Spinal Cord Segmentation using a Minimal Path in High Dimensions

To segment the spinal cord, in Chapter 2 we propose a novel method to find the globally optimal segmentation of the spinal cord using a 6D (three spatial variables, three shape variables) minimal path search. The spinal cord cross-sectional shapes are represented using principal component analysis which captures most of spinal cord's axial cross-sectional variation and partial volume effects. We propose modifications to our 6D graph to drastically

reduce the required memory and run-time to make our high dimensional minimal path optimization computationally feasible. Finally, we validate our results over five vertebrae levels of both healthy and MS clinical MR volumes (20 volumes total) and show improvements on volume agreement with expert segmentations and less user interaction when compared to a current state-of-the-art method [48]. This work was published in the 2013 International Symposium on Biomedical Imaging [34].

1.5.2 Augmenting Auto-context with Global Geometric Features for Image Segmentation

Anatomical shape variations are typically difficult to model and parametric or hand-crafted models can lead to ill-fitting segmentations. This difficulty can be addressed with a framework like auto-context [68], that learns to jointly detect and regularize a segmentation. However, since only local information is considered, mis-segmentation can still occur when a desired structure, such as the spinal cord, has few locally distinct features. High-level knowledge at a global scale (e.g. an MRI contains a single connected spinal cord) is needed to regularize these candidate segmentations.

To encode high-level knowledge, in Chapter 3 we propose to augment the auto-context framework with global geometric features extracted from the detected candidate shapes. Our classifier then learns these high-level rules and rejects falsely detected shapes. To validate our method we segment the spinal cords from 20 MRI volumes composed of patients with and without multiple sclerosis and demonstrate improvements in accuracy, speed, and manual effort required when compared to state-of-the-art methods. This work was accepted for publication in the 2013 MICCAI Workshop on Machine Learning in Medical Imaging (MLMI) [33].

1.5.3 Novel Morphological and Appearance Features for Predicting Physical Disability from MR Images in Multiple Sclerosis Patients

Physical disability in patients with multiple sclerosis is determined by functional ability and quantified with numerical scores. In vivo studies using magnetic resonance imaging (MRI) have found that these scores correlate with spinal cord atrophy (loss of tissue), where atrophy is commonly measured by spinal cord volume or cross-sectional area. However, this correlation is generally weak to moderate, and improved measures would strengthen the

utility of imaging biomarkers.

We propose novel spinal cord morphological and MRI-based appearance features in Chapter 4. Select features are used to train regression models to predict patients' physical disability scores. We validate our models using 30 MRI scans of different patients with varying levels of disability. Our results suggest that 1) regression models, trained with multiple spinal cord features, predict clinical disability better than a model based on the volume of the spinal cord alone; and 2) a non-linear, non-parametric machine learning approach can dramatically improve clinical correlation over a linear model given the same set of features. This work was accepted for publication in the 2013 MICCAI Workshop on Computational Methods and Clinical Applications for Spine Imaging [35].

1.6 Thesis Outline

The rest of this thesis is organized in the following way:

- In Chapter 2, we further examine the spinal cord segmentation problem, introduce the relevant spinal cord segmentation literature, present a novel method to segment the spinal cord using hand-crafted energy terms, a shape prior and a minimal path in high dimensions and finally present our segmentation results.
- In Chapter 3, we discuss problems with the existing segmentation approaches, including our own method described in Chapter 2, present an iterative machine learning approach with the novel addition of high-level geometric features, and demonstrate superior results to our previous method.
- In Chapter 4, we incorporate clinical data and examine how spinal cord segmentations and MRIs can be used to predict the level of physical disability in an MS patient. We extract novel morphological and appearance features and examine regression models combining these features. Our results suggest our novel features and a multiple regression model may be useful in predicting a patient's physical disability.
- Finally, in Chapter 5, we conclude by summarizing and discussing our thesis contributions and presenting some final thoughts on future work.

Chapter 2

Globally Optimal 6D Minimal Path Segmentations

As discussed in the Chapter 1, multiple sclerosis (MS) studies have shown that spinal cord atrophy contributes to a patient's physical disability (e.g. walking speed) and accurately segmenting the spinal cord increases our ability to quantify the physical atrophy of the cord. This may allow us to better monitor both the progression of the disease and the effectiveness of treatments [57]. While manual segmentations by experts are generally considered to be sufficiently accurate, they are more susceptible to operator variability and can take a considerable amount of operator time per scan [48]. This makes semi-automated methods desirable. Thus there are a number of related works on the problem of spinal cord segmentation [13, 29, 39, 45, 48, 59].

2.1 Spinal Cord Segmentation Algorithms

Schmit and Cole [59] performed segmentation using 3D seeded region growing and observed the narrowing of the spinal cord after injury in MRI scans. Horsfield et al. [29] used an active model of the cord surface where a user marked the center line of the cord on representative slices. Chen et al. [13] used a deformable atlas-based registration combined with a topology preserving classification to fully automate a crisp segmentation of the cord. Law et al. [39] segmented the cord and extracted its center-line by considering the gradient information at different scales, orientation of the detected responses, and the intensity differences between

voxels near the center of the cord. McIntosh et al. [45] segmented the cord using locally optimal 3D deformable organisms guided by a Hessian-based spinalness filter (modified from Frangi et al. [24]) designed to adapt to the varying elliptical and tubular structure of the spinal cord. This was extended by McIntosh et al. [48] where the spinal cord segmentation optimization problem was split into two sequential steps: medial path finding (based on two 2D live-wires, which can require manual corrections) followed by a medial-guided spinal crawler with cross-sectional shape fitting. Our proposed high dimensional path optimization treats these two steps in a single optimization and guarantees a global solution (from only two seed points).

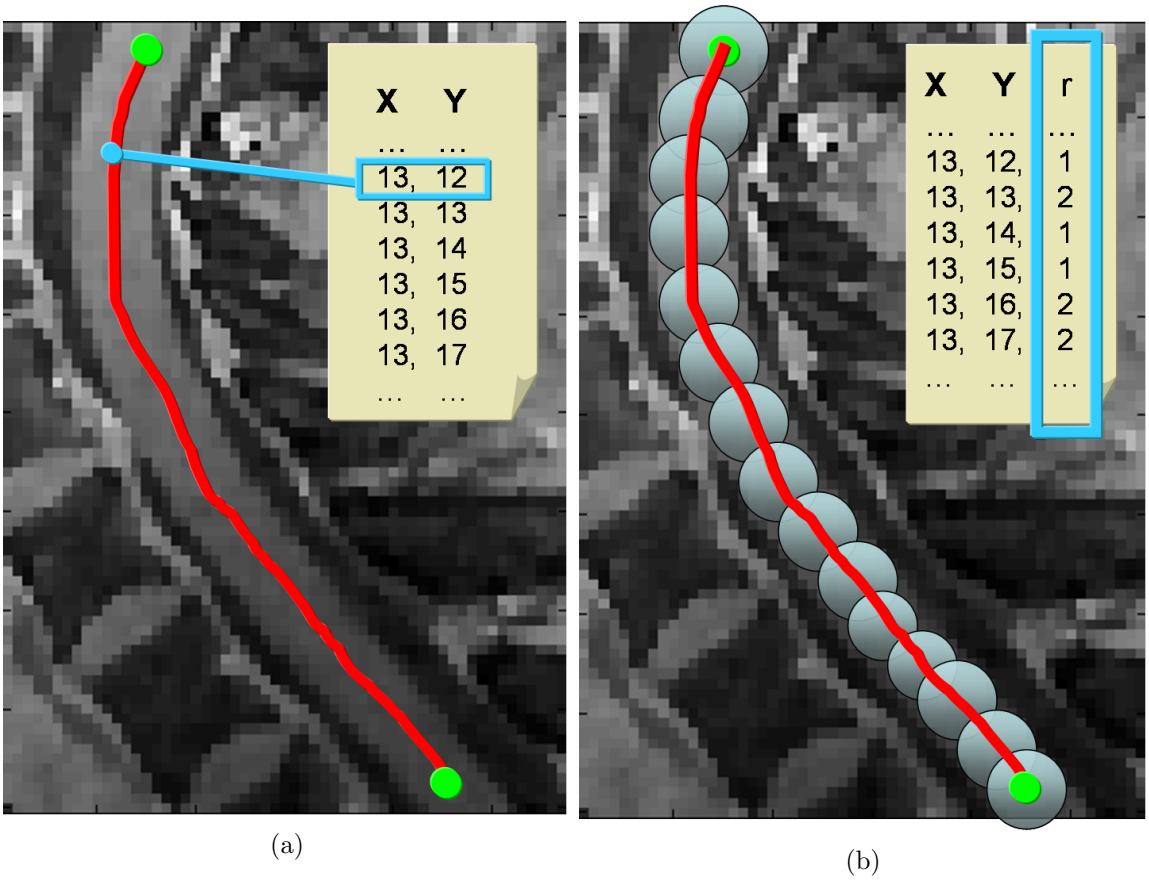
2.2 Method Introduction

We could represent the medial axis of the spinal cord in a 2D image as a list of x and y coordinates. Given a start and end seed point, if we associate a cost at each coordinate such that a lower cost represents a better choice of a medial path, we could apply a minimal path algorithm (e.g. Dijkstra's algorithm [19]) to find the 2D minimal path using a live-wire [5] based approach. This minimal path would represent the medial axis of the spinal cord (Fig. 2.1a).

To segment the cord, we could extend this method by adding a radii component. Each x, y coordinate would also have a corresponding radius measure r which would model the cord as a series of connected circles. We could extended our 2D minimal path search to a 3D search over the x, y coordinates and the radii. In this way, a 3D minimal path represents a segmentation (Fig. 2.1b).

This idea of moving to a higher dimension to find a minimal paths was introduced by Li and Yezzi [40] who applied a 4D search to find the spatial coordinates and radius of 3D tubular objects between two user entered seed points. Poon et al. [56] used a similar minimal path approach to interactively segment 2D vessels by moving to a 3D space and searching over the radii. However, these approaches only add a single dimension (the radii) which greatly limits the segmentations to what can be captured by a series of circles.

Our contribution is the extension of these methods to higher dimensional space where we demonstrate a novel method to segment the spinal cord using an A* minimal path optimizer [26]. Directly extending these methods to higher (than 4) dimensions can become



(a) (b)

Figure 2.1: A 2D sagittal slice of the spinal cord. (a) The cord's medial axis can be represented as a list of two numbers (a path) in a 2D image. (b) The cord can be modelled as a series of connected circles represented by two spatial dimensions (x, y) and a radius dimension (r) in a 2D image.

infeasible due to the high computational needs. Our approach ensures that the globally optimal minimal path can be found by automatically ignoring many suboptimal paths using a heuristic and lowers our space requirements by not requiring an explicit graph representation prior to searching. As well, we describe a probabilistic shape representation (based on principal component analysis [32]) with implicit regularization capable of capturing partial volume effects that can be extended to arbitrarily high dimensions.

Principal component analysis (PCA) has been used as an efficient method to represent high dimensional shapes by representing complex objects using only a few principal components. Cootes et al. [15] applied PCA to points on objects' segmentations to determine a statistical model of the shape changes. Cremers et al. [17] encoded an arbitrary shape where each pixel is given a probability it is within the shape and apply PCA to this model. To ensure that PCA does not leave the range of valid probability values (i.e. the unit simplex) Changizi and Hamarneh [12] proposed the Isometric Log-Ratio (ILR) transformation [20] where they performed PCA in the ILR space to ensure the probability vectors stay in the valid simplex space. Andrews et al. [2] performed PCA in this ILR space and incorporate this shape prior directly into an energy function for image segmentation. In our work we adopt the ILR-based PCA in the high dimensional minimal path search.

2.3 Methods

In this section we describe our PCA based shape representation, followed by a description of our energy function, and finally we give an overview of our minimal path optimizer. See Fig. 2.2 for a high-level outline of our proposed work-flow.

2.3.1 Probabilistic PCA based Shape Representation

We learn a model of the shape of the axial cross section of the spinal cord using PCA. This PCA based model allows us to learn a mean shape and deform it in ways that respect the main modes of variation found within the training data. By manipulating the eigenweights of the three principal components (PC), we found our model can theoretically describe a mean Jaccard similarity coefficient [31] of 0.874 and a mean area similarity of 97.18% (Section 2.4 for similarity measures) of the variation found within the training slices. With our PCA based shape model, we represent a single axial slice of the cord using the following six dimensions: the x, y, z spatial coordinates describing the centroid of the cross section,

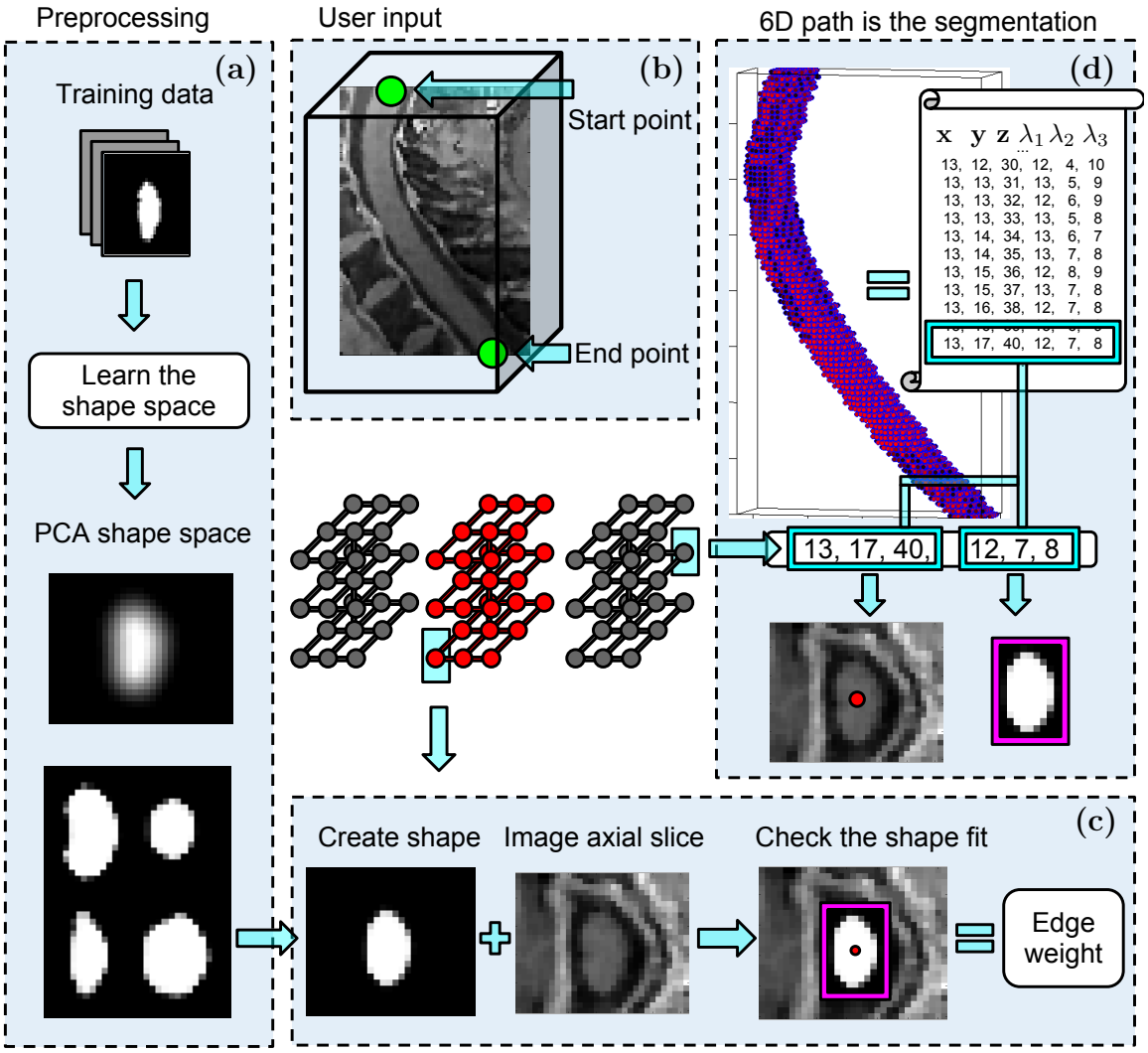


Figure 2.2: The proposed work-flow of our method. (a) As a preprocessing step, we apply PCA to the ground truth data to learn the shape space of the axial slices. (b) The user enters a start and end seed point. (c) Beginning from the start seed point, we use our energy function to determine how well the PCA shapes fit to the image data. (d) A single shape is represented by six numbers, where three spatial coordinates (x, y, z) and three PC weights $(\lambda_1, \lambda_2, \lambda_3)$ are used to describe the shape. The stack of these 2D segmentations form our final 3D segmentation, which is optimized by finding a minimal path in the 6D search space between the start and end seed points.

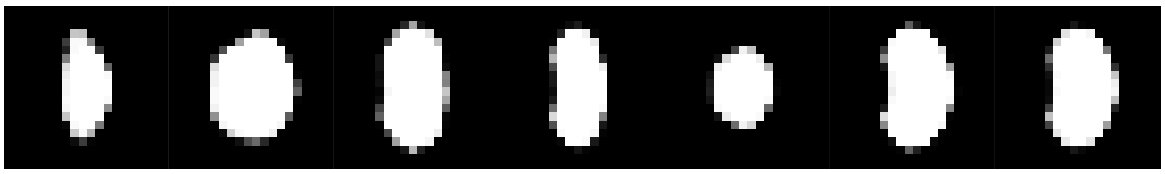


Figure 2.3: Sample spinal cord cross-sectional shapes generated by randomly sampling the PC weights. Note the probabilistic pixels and the non-ellipsoidal shapes.

and three PC weight values for describing its shape. We restrict the PC weights to only vary by ± 3 standard deviations.

To capture the partial volume effects, our shape is embedded into an image such that each pixel has a value associated with it between 0 and 1. This value represents the probability that a random point inside a single pixel contains the spinal cord. Hence a value of 0 would signify the pixel does not contain the spinal cord and a value of 1 represents a pixel that is fully part of the spinal cord.

Since we adopted a probabilistic representation, we need to consider that standard PCA is not constrained to the geometry of the unit simplex (space of non-negative vectors with unity sum) and can produce shapes that are invalid (outside the simplex) [12]. To stay within the simplex, we map our data to the ILR space before applying PCA and then map the resulting shape back to our probabilistic shape space [2]. To convert to ILR space, the following ILR transformation is derived for the two label case,

$$q = \mathrm{ILR}(\boldsymbol{p}, \boldsymbol{e}) = \ln\frac{p_1}{g(\boldsymbol{p})}\ln\frac{e_1}{g(\boldsymbol{e})} + \ln\frac{p_2}{g(\boldsymbol{p})}\ln\frac{e_2}{g(\boldsymbol{e})} \tag{2.1}$$

where, $\boldsymbol{p} = [p_1, p_2]$ is a vector of our probabilistic values; $\boldsymbol{e} = [0.804, 0.196]$ is the Aitchison-orthonormal basis in the simplex; and $g(\boldsymbol{x}) = \sqrt{x_1 x_2}$ is the geometric mean. To convert back to our probability space, we calculate,

$$\boldsymbol{p} = \mathrm{ILR}^{-1}(q, \boldsymbol{e}) = (e_1^q + e_2^q)^{-1}\left[e_1^q, e_2^q\right]. \tag{2.2}$$

Given our learned shape space, we can apply varying PC weights to it to deform our shape as can be seen in Fig. 2.3.

2.3.2 Energy Function

Our energy function is designed to model two characteristics of the spinal cord in T1 weighted MRI: 1) the cord is a lighter object usually surrounded by an outer dark ring

of cerebrospinal fluid, E_{∇}, and 2) the intensities found in the spinal cord should exhibit a similar degree of homogeneity as found in the training data, E_{σ}. We define E our energy function as

$$E = \alpha E_{\nabla} + \beta E_{\sigma} \tag{2.3}$$

where α, β are weights for each energy term (both set to 1 for our experiments). Each of these terms are data terms as the regularization is built directly into our shape representation and optimizer; we constrain the PC weights (i.e. shape) to only vary by one standard deviation across slices, and enforce the spatial regularization through graph connectivity. Note that in our formulation, we assign the same energy to all shapes in the allowable shape domain, regardless of how different they are (how far they are) from the mean shape.

In E_{∇}, our goal is to ensure the gradient vectors on the object's boundary point towards the center of the cord:

$$E_{\nabla} = \frac{\int_{\Omega} \left(\left(f\left(\nabla P(x,y)\right) \circ \nabla I(x,y) \right) \cdot R(x,y) \right) dx\ dy}{\int_{\Omega} \left| f\left(\nabla P(x,y)\right) \right| dx\ dy} \tag{2.4}$$

where $(x, y) \in \mathbb{R}^2$ is a point in a 2D image domain Ω for a given plane along the z-axis; P is the probability of each pixel belonging to the cord (Fig. 2.3); I represents our image; R is composed of vectors pointing towards the centroid of the shape found in P; f is a Heaviside step function to only allow for strong edges and to treat all edges equally; and $\circ$ and $\cdot$ denote element wise multiplication and the dot product, respectively.

Our second energy term, E_{σ}, assigns low energy to shapes that contain pixel intensities whose standard deviation $\sigma_p = \text{std}(I(x,y); (p\,(x,y) > \epsilon))$ in a 2D slice agrees with standard deviations, $\boldsymbol{\sigma}$, of spinal cord intensities collected across all slices in the 3D training data. We define $\bar{\sigma}$ as the mean learned standard deviation of intensities (i.e. mean($\boldsymbol{\sigma}$)); std($\boldsymbol{\sigma}$) as the standard deviation of the learned standard deviation of intensities, $\boldsymbol{\sigma}$, computed over the 3D volume. We set $\sigma_{lower} = \bar{\sigma} - 3 * \text{std}(\boldsymbol{\sigma})$, and $\sigma_{upper} = \bar{\sigma} + 3 * \text{std}(\boldsymbol{\sigma})$. E_{σ} penalizes shapes whose σ_p exceeds three standard deviations from the learned $\bar{\sigma}$, calculated as:

$$E_{\sigma}(\sigma_p, \boldsymbol{\sigma}) = \begin{cases} (\sigma_{lower} - \sigma_p), & \text{if } (\sigma_p < \sigma_{lower}) \\ (\sigma_p - \sigma_{upper}), & \text{if } (\sigma_p > \sigma_{upper}) \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

2.3.3 Minimal Path Optimizer

We optimize our energy function using an A* minimal path search. A standard approach would be to use Dijkstra's algorithm [19] and apply it to high dimensions (initially we used the N-D Dijkstra's algorithm provided by [74]). However, this approach is problematic since it (and most traditional minimal path methods) requires as input to the minimal path algorithm a graph composed of nodes and edges. The memory required to explicitly encode all the edges of our graph explodes due to the high connectivity of 6D graphs. For our 6D volume, a *single* node connected to all its immediate neighbours (26-connected in 3D space) would require $3^6 = 729$ edges. We found that even with modifications to the existing N-D Dijkstra's code and underlying data structures [74], we were unable to lower our memory requirements to make our method computationally feasible (e.g. running our method on a single cropped volume of $25 \times 20 \times 80$ could require over 70 GBs). As well, a further drawback is the considerable run-time needed to create such a graph. Fig 2.4 illustrates the magnitude of a 6D graph. In addition, an advantage that A* has over Dijkstra's algorithm, is the ability to define heuristics to speed up the run-time (our run-time is between 1 to 5 hours) while maintaining globally optimality.

To reduce both the run-time needed to create the entire graph and the memory required to store the edges of the graph, we observed that the structure of a volume allows us to infer, on-the-fly during run-time, the connectivity of a voxel. Thus rather than explicitly storing all the edges as input, we modify the minimal path algorithm to index into our volume and determine the appropriate neighbours at run-time. This removes the need to encode the edges prior to running our minimal path method. As we now only need to keep a subset of the edge values in memory (those in the minimal path "open set" queue), our memory requirements drop drastically which allows us to search over larger volumes (e.g. the 70 GB dropped to 4 GB).

We "prune" certain edges of our graph to create a regularization prior. We use the prior knowledge that the spinal cord in the MRI volume extends superior to inferior i.e. along the z dimension. We thus remove any edge in between slices that does not transition in the z dimension towards the end (inferior) seed point. As well, we discretize the possible PC weights to use 7 possible multiples of the standard deviation values $\{0, \pm 1, \pm 2, \pm 3\}$.

We implement an efficient low-memory A* search to find the globally optimal path between two user entered seed points where the returned minimal path contains the 6D

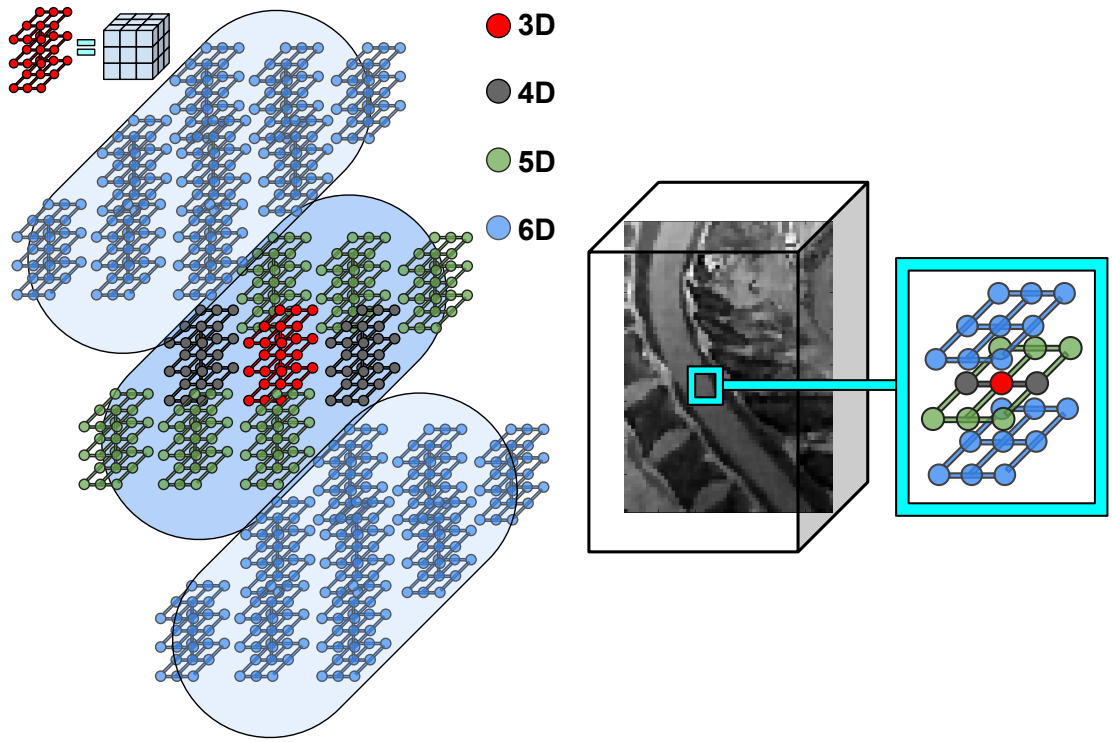
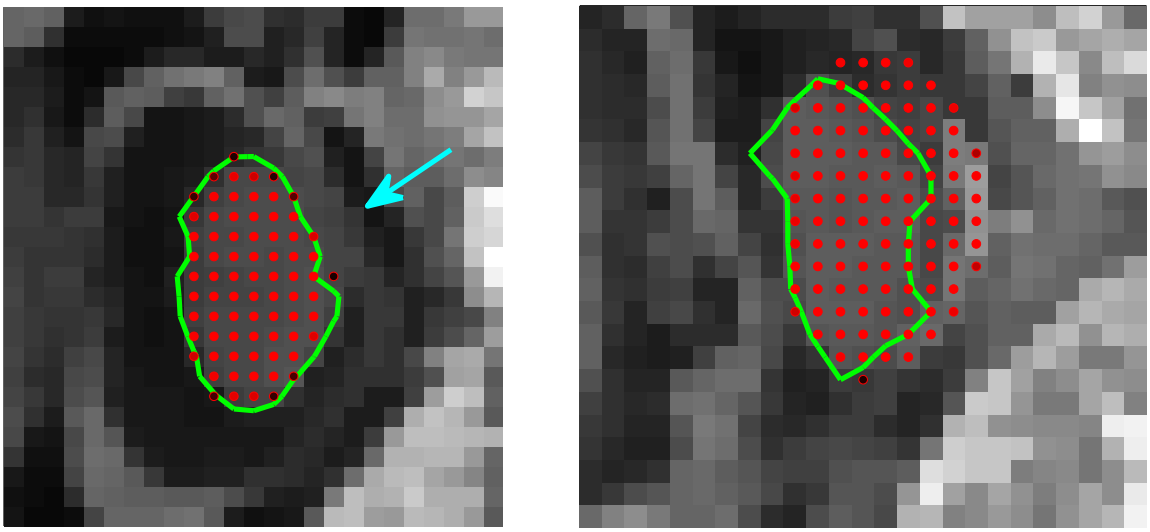


Figure 2.4: (Left) To get a sense of the exponential increase in the size of a 6D graph, we illustrate the 6D search space for a $3 \times 3 \times 3$ volume. (Right) An alternative visualization where each 3D voxel has three shape dimensions associated with it. Note that many edges have been removed for visualization purposes.



(a) (b)

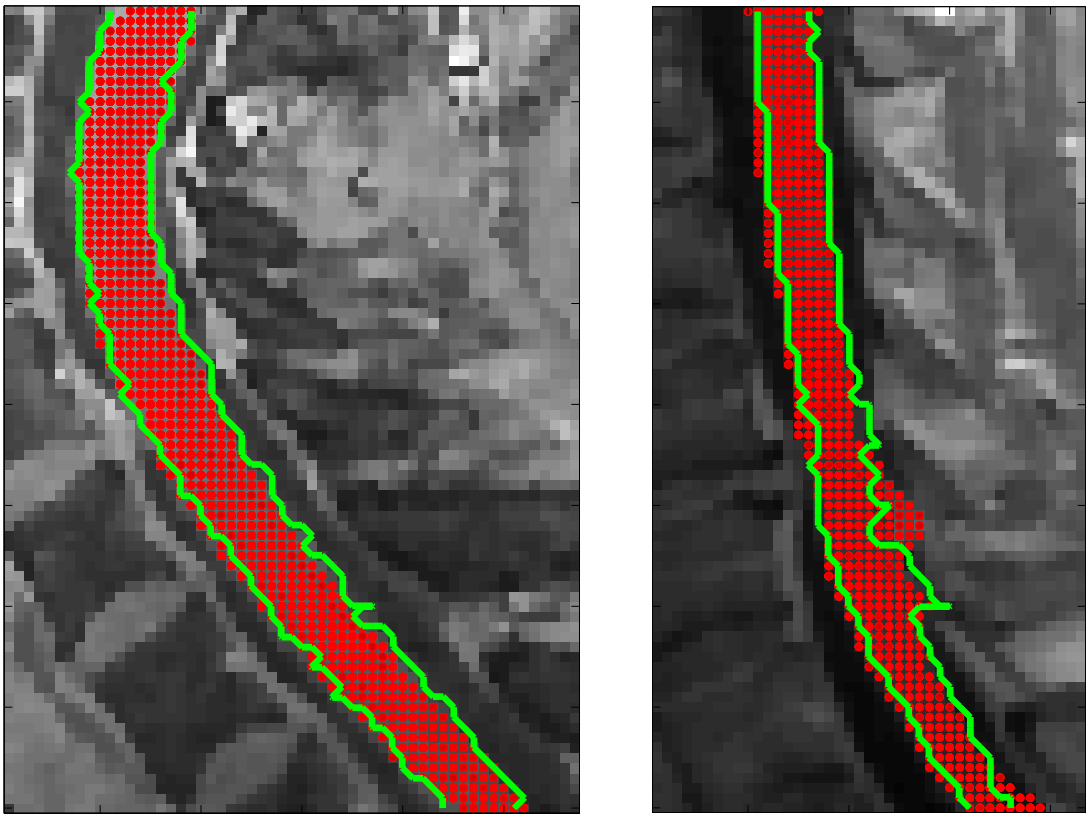
Figure 2.5: Segmentation results where the green contour represents the ground truth and the red dots represent our segmentation method. The darker red dots on the boundary represent lower probability of the spinal cord (*visible in the electronic version*). (a) Axial slice with misleading image data (*arrow*) that our method can successfully segment. (b) Axial slice where our method latches onto the wrong border due to misleading gradient information.

coordinates that represent the globally optimal segmentation.

2.4 Results

We validate our method over 20 MRI scans composed of 10 healthy and 10 MS patients scanned with a 3.0T scanner and 1.5T scanner respectively (scans were from different studies) with a voxel size of $0.9760 \times 0.9760 \times 1$ mm. Each scan was segmented by an expert and, given its use in previous studies [48], we consider this a reliable expert segmentation. We validate our method over a challenging area of the cord (C3 - C7) of 80 slices.

To seed our method, we simulate the user clicking seed points by extracting the approximate center of the expert segmentation for the first and last expertly segmented slice to get the first three spatial dimensions. We use the mean shape of our spinal cord PCA model (we build a separate MS and non-MS PCA model with the testing volume omitted) as the three PC weight dimensions (i.e. zero weights). While this does not give an optimal



(a) (b)

Figure 2.6: Segmentation results where the green contour represents the ground truth and the red dots represents our segmentation method. (a) Sagittal MR slice where our segmentation works reasonable well. (b) Sagittal MR slice where our segmentation latches onto incorrect borders. Note the abnormal protrusion that also appear in the border of the ground truth. This is an example of where occasionally, the ground truth appears to latch onto incorrect borders, demonstrating the difficulty in producing accurate segmentations even with heavy amounts of user interaction.

segmentation for the first and the last slice, our method quickly converged to a reasonable segmentation after one or two slices. A user could get around this limitation by placing the seed a few slices before and after the desired locations, or specifying initial major and minor axis lengths. We use our start and end seed point to roughly crop our volume around the cord to help eliminate voxels from our search space that cannot be part of the spinal cord.

Given the small structure of the spinal cord it is important to consider the contributions of partial volume effects (PVE) on the segmentations to ensure accuracy (darker red dots in Fig. 2.5). The Jaccard similarity index is a commonly used measure to determine the similarity between shapes and is defined as $J(A,G) = \frac{|A \cap G|}{|A \cup G|}$. We modify Jaccard to consider PVE and decrease the score as the PVE for each pixel increases in difference between each shape. We calculate this as,

$$J_{\mathrm{PVE}}(A,G) = \frac{\sum_x \sum_y \min(A(x,y), G(x,y))}{\sum_x \sum_y \max(A(x,y), G(x,y))} \tag{2.6}$$

where A and G are the automated and expert (ground truth) shapes embedded into the same sized image. The "min" function represents a probabilistic intersection and the "max" function represents a probabilistic union. We include the crisp version of Jaccard (round PVE) in our results denoted as J_{MAP}.

To determine whether the automated method is a useful surrogate of the manual method for detecting global cord atrophy, we also compare the computed volumes between the automated and expert segmentations regardless of their overlap. As the cross-sectional area (CSA) has been shown to be a good indicator of the progression of MS, a high agreement between an automated and expert segmentation would be a good indication of the clinical utility of a particular method. To measure this, we use a similarity ratio, Vol_{S}, as described in [48]. The Vol_{S} is given as $1 - \min(|(A/G) - 1|, 1)$ which returns a 0% accuracy if the automated volume overestimates the expert volume by more than 200%.

In Table 2.1 we see how our automated segmentation compares with the expert segmentation. As might be expected, our method performs better overall on the healthy patients than the MS patients due to the increased irregularity of the spinal cord as MS progresses and the change from a 3T to a 1.5T scanner. Our results compare favourably with other similar methods and requires less user input. In [48], which requires some user correction of the live-wire path, they report a mean volume similarity Vol_{S} of 94.31% for MS and 91.60% for healthy patients. Noticeably, for healthy patients they report a min similarity of 77.33%,

Table 2.1: Similarity in shape overlap (J_{PVE} = probabilistic and J_{MAP} = binary Jaccard index) and computed area (Vol_{S}) between automated and expert segmentations for multiple sclerosis patients (*left columns*), and healthy (non-MS) patients (*right columns*).

	J_{PVE}	Vol_{S}	J_{MAP}	J_{PVE}	Vol_{S}	J_{MAP}
Mean	0.784	94.33	0.789	0.832	96.88	0.855
Median	0.797	95.15	0.800	0.848	97.77	0.869
Min	0.451	88.72	0.473	0.379	90.32	0.394
Max	0.906	99.88	0.949	0.915	99.81	0.989
Std.Dev	0.066	3.30	0.076	0.069	2.94	0.083

compared to our reported min similarity of 90.32% which suggests our method is more capable of capturing the cord's true volume. Segmentation results highlighting typical examples where our method works well and where the method gets mislead due to misleading image information are shown in Fig. 2.5 and Fig. 2.6.

2.5 Conclusion

We proposed a method to segment the spinal cord by finding a minimal path in 6D between two user entered seed points. This minimal path represents the globally optimal segmentation of the spinal cord under the appropriate assumptions (e.g. discrete PCA weights, gradient and standard deviation term, etc.). Although we used a 6D representation, this method could be extended to higher dimensions but at a cost in both memory requirements and computation time. We address some ways to reduce this in the form of our optimized A* search that take advantage of the volume structure and edge pruning to drastically reduce the needed memory. Future work would look at further space and run-time optimizations, study reproducibility and sensitivity with respect to seed point selection, and correlate the automated segmentations to the progression of MS.

The accuracy of this method could be improved by modifying the energy function of Equation (2.3) with more descriptive hand-crafted energy terms or a machine learning based approach to detect the cord. As well, using only the three principal components with a discrete search space and restricted graph connectivity limits the available shapes that can be created, preventing us from capturing the full range of axial shapes that appear in our segmentations. We discuss these limitations further in the following chapter and propose a method to improve both detection and regularization of the spinal cord.

Chapter 3

Global Geometric Features for Auto-context

3.1 Introduction

In our previous chapter, to segment the spinal cord we represented the axial shape variations of the spinal cord using probabilistic principal component analysis (PCA) and found the globally optimal path in 6D (three spatial and three principal components weights) between two user specified seed points [34]. However, this approach is limited (e.g. Fig. 2.5 and Fig. 2.6) by the hand-crafted energy functions used to detect the spinal cord, and the restrictions (e.g. discrete PCA weights) on the shape space. These limitations motivate our development of the machine learning approach described in this chapter.

Our method described in Chapter 2 and many of the other previously mentioned spinal segmentation methods (Sec. 2.1), rely on a balance between appearance (the information in an image) and regularization (the expected properties of the segmentation) expressed in the form of an explicit energy functional to minimize.

In contrast to this paradigm, Auto-context, proposed by Tu [67], is a general iterative learning framework used for segmentation that jointly learns the appearance and regularization distributions where the predicted class labels (the context) of the previous iteration are used as input to the current iteration. Auto-context was shown to improve segmentation results [68]; however, if the surrounding local context is incorrect it may propagate and reenforce an incorrect segmentation. This was addressed in the recent work by Kontschieder

et al. [36] who used a geodesic distance transform in an auto-context based segmentation approach to incorporate long range spatial context. Similarly, we extend auto-context to learn high-level problem specific information (global context), but rather than considering the geodesic distance, our probability mask represents distinct candidate shapes and encodes their relations to each other to encourage a single connected spinal cord.

Specifically, we propose to extract geometric features (e.g. volume) from the candidate shapes created in each iteration of auto-context. We define a shape as a distinct region of connected components composed of the same class label. Features from the candidate shapes are extracted and compared with each other to give global information about the other candidate shapes and are included in the auto-context framework. Thus the class label for a specific voxel is conditioned not only on its appearance and surrounding labels, but also on the geometric features of the shape it belongs to relative to other candidate shapes. We demonstrate that augmenting auto-context with global geometric context improves the original auto-context algorithm, and results in superior spinal cord segmentations when compared to the more traditional approach using hand-crafted gradient and intensity-based terms regularized by a PCA shape prior (Eq. 2.3) described in Chapter 2.

3.2 Methods

This section introduces auto-context, describes our global geometric features, explains our auto-context set-up, and outlines the training and testing processes.

3.2.1 Image Segmentation and Auto-context

We can think of image segmentation as a voxel labelling problem, where given an image $X = (x_1, \ldots, x_n)$ composed of n voxels, we want to find a set of labels such that each voxel i is assigned a corresponding label, $Y = (y_1, \ldots, y_n)$. The label y_i can take on one of k possible values. Our objective is to find the optimal configuration Y^* out of the possible segmentations Y such that it maximizes the probability given the observed image, $Y^* = \text{argmax}_Y\, p(Y|X)$.

One approach to find Y^* is to apply Bayes' rule [10] with a fixed $p(X)$ to give $p(Y|X) \propto p(X|Y)p(Y)$, where $p(X|Y)$ represents the likelihood of the image given a segmentation, and $p(Y)$ is the probability of that segmentation occurring. Hand-crafted data-terms and/or

parametric models with restrictions on complexity and built-in assumptions (e.g. Gaussian) are often used [34, 45, 48].

Another approach is to directly model $p(Y|X)$. If we assume that the labels $y_i, i = 1, \ldots, n$ are independent and are conditioned only on a small patch of image voxels centred around the ith voxel, denoted as $N(i)$, then this can be modelled by $p(y_i|X_{N(i)})$, where X_i returns the intensities at voxel i. A discriminative model (e.g. decision forest [18]) can be used to learn this distribution; however, it does not consider the class labels of surrounding voxels. One way to model the interdependence of neighbouring labels is with conditional random fields [38] where the class label y_i is dependent on a neighbouring class label y_j [64]. While this more closely approximates the true $p(Y|X)$, it still makes the assumption that each class label y_i is only dependent on a very small neighbourhood.

The auto-context model seeks to condition over a larger area of surrounding class labels to provide more "context" [68]. This is accomplished by introducing an iterative time-step t and training a series of classifiers using the discriminative probability (the classification confidence) map of the previous classification M^{t-1} where $M^t = (\boldsymbol{m}_1^t, \ldots, \boldsymbol{m}_n^t)$. Each vector $\boldsymbol{m}_i^t$ represents the probabilities of voxel i belonging to one of the k possible class labels, $\boldsymbol{m}_i = [p(y_i = 1), \ldots, p(y_i = k)]$. The initial class probability map M^0 is set to have uniform values. At time t, a classifier is trained to predict the true class label y_i given the image patch $X_{N(i)}$ and the context information $M^{t-1}(i)$, where M is centred at voxel i. Once the classifier is trained, the new probability map M^t is used in the next iteration $(t + 1)$ and the algorithm repeats until M converges. The final output is a series of learned probability distributions,

$$m_i^t = p^t(y_i|X_{N(i)}, M^{t-1}(i)). \tag{3.1}$$

In testing, a novel image has the same features extracted and goes through the iterative classification process using the learned probability distributions p^t.

This formulation does not capture high-level information about the entire shape that the pixel belongs to nor does it consider the interaction among other candidate shapes. If we simply increase the size of $N(i)$, we increase the dimensionality of the feature space which can decrease the efficiency of our classifier. Thus this formulation is not well suited to capture high-level knowledge such as that the spinal cord is a single connected structure of a particular size.

3.2.2 Global Geometric Features

We propose to augment auto-context with global geometric features. At iteration t, instead of only using the previous probability map M^{t-1}, we also extract features from the candidate shapes found within M^{t-1} and compare these features to each other to capture global information. More formally, we define a function $C(M, i)$ that takes the maximum a posteriori (MAP) class label of M to form a candidate shape S_q based on the regions of connected component with the same class label, $S_q = C(M, i)$. This divides M into Q distinct non-overlapping connected regions with the same class label, where each distinct region represents a candidate shape S_q (Fig. 3.1c,g). The shape S_q is composed of indices into the probability map M and the index for voxel i belongs to a single shape, $i \in S_q$.

We define and utilize a single global geometric feature that encourages our method to detect only a single connected spinal cord. For voxel i we extract a feature $f(S_q, M)$ from the shape that i belongs to and compute the ratio between the current and the largest shape feature from the Q candidate shapes,

$$\phi(M, i) = \frac{f(C(M, i), M)}{\max\left(f(S_1, M), \ldots, f(S_Q, M)\right)}. \tag{3.2}$$

The feature we extract, $f(S_q, M)$, gives us an indication of the size and label confidence of the component being considered,

$$f(S_q, M) = \sum_{j \in S_q} M(j) \tag{3.3}$$

where the iterator j sums the probabilities belonging to shape S_q. If $\phi(M, i)$ returns a value of 1, then this indicates that voxel i belongs to the largest probable shape out of all the candidate shapes. We augment the previous auto-context model (3.1) with our global geometric feature ϕ,

$$m_i^t = p^t(y_i | X_{N(i)}, M_{N(i)}^{t-1}, \phi(M^{t-1}, i)). \tag{3.4}$$

This model considers *intensity information*, *local context*, and the proposed *global features* about the candidate shapes and the relations between them.

3.2.3 Auto-context Setup and Feature Design

The original auto-context work used probabilistic boosting trees as the discriminative classifier, but other classifiers can be used [68]. We chose to use decision forests for our discriminative model due to their ability to generalize well to unseen data, handle both classification

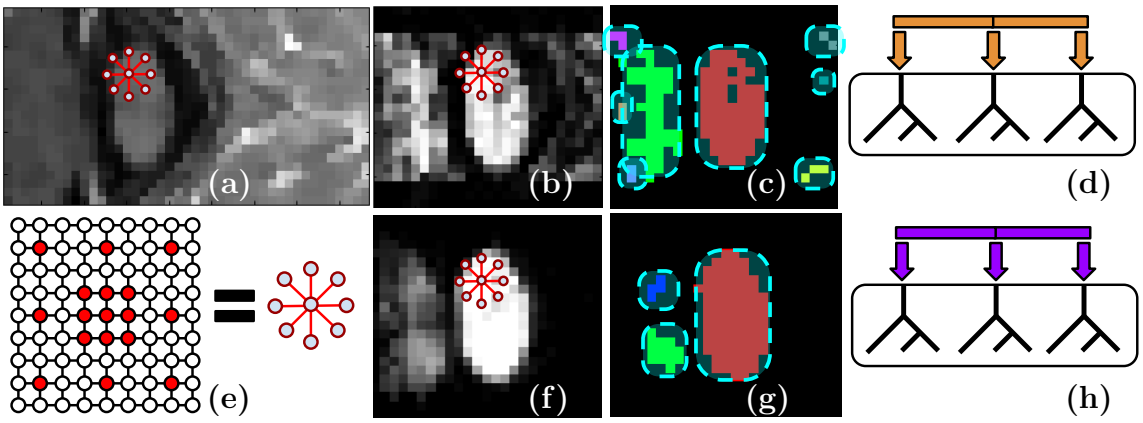


Figure 3.1: An outline of our method: (a) The MRI data X. (b) The probability map M^t. (c) The MAP estimate of M produces distinct shapes (*dotted outline*) whose features $\phi(M, i)$ can be extracted. (d) Decision forest trained on *a,b,c*. (e) The neighbourhood $N(i)$ in 2D. (f) The probability map M^{t+1} produced by *d*. (g) Shapes computed from *f*. (h) Decision forest trained on *a,f,g*.

and regression, learn a distribution, and provide a probabilistic output [18]. This probabilistic score $\boldsymbol{m}_i$ is particularly important because it gives us an indication of the probability of being a member of the class by computing the percentage of observations of this class in a tree leaf averaged over all trees.

The neighbourhood of the ith voxel, $N(i)$, is computed using a simple radial pattern where we sample those voxels that are the immediate neighbours (8 in 2D, 26 in 3D) of voxel i and those that are three voxels along the ray away from voxel i (Fig. 3.1e). This densely samples points close to our ith voxel and sparsely samples voxels further away which helps keep the size of the feature vector relatively small to allow for faster run-times. We use a 52-neighbour 3D version of the neighbourhood shown in Fig. 3.1e. To compute image appearance features we directly sample the neighbourhood intensities. In order to have an invariance to a shift in intensities and to capture the polarity (e.g. dark-to-bright transitions), we divide the intensity of neighbourhood voxels by the intensity of the voxel of interest, $X_{N(i)} = \{\frac{x_1}{x_i}, \ldots, \frac{x_{52}}{x_i}\}$.

3.2.4 Training and Testing

Since the two patient groups should not be modelled as samples from a single distribution, we separate the multiple sclerosis (MS) from the non-MS patients to train two different

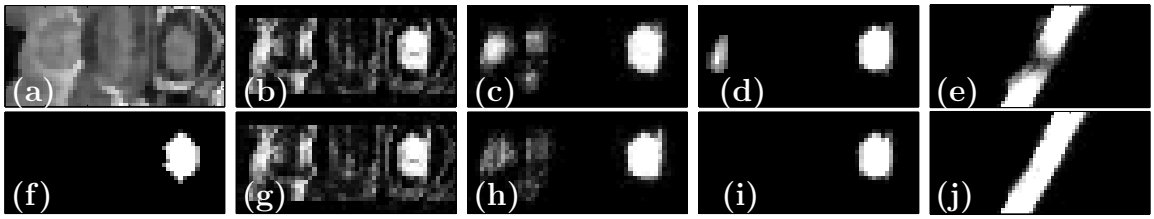


Figure 3.2: Segmentation without global features *row 1*, and with global features *row 2*. (a) An axial slice of the cropped spinal cord. (f) The ground truth. (b,g) First iteration over-detects the cord. (c,h) Second iteration begins to regularize the segmentation. (d) Final fuzzy border segmentation with a false positive. (i) False positive removed by the global features. (e) Sagittal plane where weak appearance information splits the cord. (j) Corrected with global features.

classifiers specifically on the variability found in the two groups. We employ leave-one-out testing.

Our auto-context model is trained based on (3.4). Our training segmentations have a value of 1 inside the cord, 0 outside the cord, and a partial volume estimate for the cord's boundary. Each decision forest p^t is trained on a subset of this training data (2000 voxels from each volume) split between cord and background samples with the fuzzy border voxels omitted and samples of false positives explicitly included if they exist. To reduce overfitting, for each training volume V, a separate decision forest that does not include the data from V is trained and used to predict the class membership of V. We repeat this process 5 times as we found that the results generally stabilized without further training (less than 0.01 difference in the mean Jaccard similarity index between consecutive iterations). The training produces a set of 5 trained decision forests, $\{p^1, \ldots, p^5\}$.

As our ground truth segmentations have a fuzzy border based on partial volume estimation not modelled by the decision forests, we train a separate regression forest p^r to capture a two-pixel thick partial volume border using the same intensity and probability features. It can take up to 6 hrs to fully train our auto-context model on a machine using 4 cores.

In the testing phase, a novel volume is classified by all the trained decision forests, $\{p^1, \ldots, p^5\}$, in series. We take the maximum a posterior estimate of our final classified volume M^5 and use the regression forest trained on border partial volumes p^r to produce our final segmentation (Fig. 3.2i,j).

3.3 Results

We validate our method using the same 20 MRI scans from Chapter 2, composed of 10 MS patients from a 1.5T scanner and 10 healthy patients from a 3.0T scanner with a voxel size of 0.976 $\times$ 0.976 $\times$ 1.000 mm. The same expert segmentations were used as the ground truth data as in Chapter 2. Briefly, they were produced using an in-house method similar to that by Tench et al. [66]. We segment over the same vertebra, C3 - C7 spanning a total of 80 slices.

To capture the partial volume effects (PVE) [66], we use a probabilistic version of the Jaccard index defined earlier in Equation (2.3). To highlight our methods improved ability to remove false positives, we compute the Hausdorff distance [30] which measures the furthest distance between the closest points in the two segmentations,

$$H = \max_{a \in A} \left(\min_{g \in G} D(a, g) \right) \tag{3.5}$$

where $D(a, g)$ computes the 3D Euclidean distance between points a, g. To measure the similarity between the automated and manual segmentation volumes, we calculate,

$$Vol_{\mathrm{S}} = 1 - \min(|(|A|_1/|G|_1) - 1|, 1) \tag{3.6}$$

which returns a 0% accuracy if the automated volume overestimates the expert volume by more than 200% [48]. To indicate if our automated method is a useful consistent substitute to the manual method for computing spinal cord volume, we compute the Pearson's correlation coefficient [72] between the volumes of the automated and manual segmentations.

We perform three experiments to validate our method. Our **first experiment**, compares our auto-context augmented with the global features method against the segmentation method we developed in Chapter 2 [34]. This previous work modelled the cord using hand-crafted gradient and intensity-based terms and a probabilistic PCA shape prior. Our new results compare favourably, with a mean J_{PVE} of 0.878 (previous 0.784) for MS and 0.928 (previous 0.832) for non-MS scans, validated over the same dataset.

Our **second experiment** is done to better understand our improvements and to highlight the problems with restricted parametric models and hand-crafted terms. To examine the effect that the restricted parametric PCA model has on the segmentation, we pass only the *ground truth* data (i.e. the cord is perfectly detected) to be regularized by the PCA model and A* optimizer of Chapter 2 [34]. We find that even with the ground truth data,

Table 3.1: GT_{PCA} uses the *ground truth* as input to the method of Chapter 2 [34] to demonstrate the best the restricted PCA model can achieve. $PCA_{\nabla\sigma}$ represents the results using Chapter 2's approach copied here from Table 2.1 for comparison. PCA_{M} uses the spinal cord probability map M^1 as input to the PCA-based method of Chapter 2 [34]. XM is auto-context with the image and probability map features. $XM\phi$ is auto-context with the image, probability and global geometric feature. We compute the mean values of the probabilistic Jaccard index ($\overline{J_{\mathrm{PVE}}}$), area similarity ($\overline{Vol_{\mathrm{S}}}$) and Hausdorff distance ($\overline{H}$) in voxels between our automated and ground segmentations over ten MS patients (*left columns*) and ten healthy (non-MS) patients (*right columns*). The Pearson's correlation coefficients (r) and p-values (p) between the automated and the ground truth segmentation volumes are measured.

Case	$\overline{J_{\mathrm{PVE}}}$	$\overline{Vol_{\mathrm{S}}}$	$\overline{H}$	r	p	$\overline{J_{\mathrm{PVE}}}$	$\overline{Vol_{\mathrm{S}}}$	$\overline{H}$	r	p
GT_{PCA}	0.853	98.36	1.96	0.997	3×10^{-10}	0.881	98.97	1.57	0.993	9×10^{-9}
$PCA_{\nabla\sigma}$	0.784	94.33	-	-	-	0.832	96.88	-	-	-
PCA_{M}	0.818	**96.38**	2.42	0.931	9×10^{-5}	0.847	93.65	2.04	0.896	4×10^{-4}
XM	0.856	94.50	15.35	0.932	9×10^{-5}	0.916	97.74	4.41	0.939	6×10^{-5}
$XM\phi$	**0.878**	96.32	**2.29**	**0.972**	3×10^{-6}	**0.928**	**98.87**	**1.82**	**0.991**	3×10^{-8}

the restricted PCA model produces less accurate J_{PVE} results when compared to our auto-context model (Table 3.1 - row GT_{PCA}). To show the limitations of hand-crafted terms, we replace the gradient and intensity-based terms (Equation 2.3) of Chapter 2 [34] with the spinal cord probability map M^1 (trained only on intensity features) as input to be regularized by the PCA model. This improves the segmentation results (Table 3.1 - row PCA_{M}) over what was originally reported in Chapter 2 [34] indicating that the trained classifier is better at detecting the cord.

Our **third experiment** compares auto-context with and without global features. To compare to the approach of Tu and Bai [68], we initially omit the global feature and rely only on the image X and probability maps M as input to the auto-context model to segment the spinal cord (Table 3.1 - row XM). We then reintroduce the global feature $\phi(M, i)$ and show improved segmentation results (Table 3.1 - row $XM\phi$). This demonstrates that it is useful to augment auto-context with the global shape feature (i.e. improvement over [68]).

We note that, once trained, our unoptimized implementation takes under 10 minutes to run for a novel cropped volume while in Chapter 2 [34] we reported run-times between 1 and 5 hours for the same sized volumes. Also, while our method was tested on a cropped volume (Fig. 3.1a), it did not require any further user-input in contrast to [29, 48], and it

was capable of capturing the PVE not modelled in [13].

3.4 Conclusion

We have proposed to augment auto-context with global geometric features that can capture high-level information and relationships between the candidate shapes found within a segmentation. The auto-context approach jointly detects and regularizes a segmentation which allows for a flexible shape space capable of capturing subtle irregularities in the spinal cord. We have demonstrated improvements to accuracy and a lower running time when compared to our previous approach in Chapter 2. We leave further comparison of these two methods for the final Chapter 5 in Section 5.1. Future work would validate our method over a larger spinal cord dataset with more diverse clinical parameters and investigate the development of other global geometric features.

Chapter 4

Predicting MS Disability using MRI

4.1 Introduction

As discussed in Chapter 1, Multiple sclerosis (MS) studies have found that a patient's physical disability correlates with spinal cord atrophy [8, 41, 42, 57, 69]. Measuring spinal cord atrophy is potentially a useful biomarker for monitoring the progression of diseases or the effectiveness of therapies [57]. Spinal cord atrophy is defined as a loss of tissue and commonly measured by cross-sectional area (CSA) or spinal cord volume [41, 42, 57]. To quantify the CSA, user-guided computer software is often used to assist in delineating the spinal cord from a 3D MRI (e.g. using one of our previous chapters' methods [33, 34] or several other recently developed approaches [29, 45, 48, 66]). The segmented cord's volume or averaged CSA is computed and correlated with the patient's clinical disability score.

To quantify the clinical disability of a patient with MS, clinicians commonly rely on the Expanded Disability Status Scale (EDSS) [37] which assigns the patient a number between zero (a normal neurological exam) and ten (death from MS). Although commonly used, the EDSS score suffers from reproducibility issues, focuses largely on a patient's ambulatory impairment, and is restricted to an ordinal scale. This motivated the development of the Multiple Sclerosis Functional Composite (MSFC) score [22], which we discuss in Section 4.2.5.

While the CSA of the spinal cord has been shown to correlate with clinical score, this

correlation is generally moderate with some studies failing to show the expected reduction in CSA [43]. This may be because a reduction in cord size is only one global aspect of atrophy, and few other features that capture more subtle aspects have been explored. Schnabel et al. [60] explored local and global shape measurements across scales and concluded that the spinal cord shape should be measured across a range of scales. In conventional and diffusion tensor (DT) MRIs, Benedetti et al. [8] identified the brain T2 lesion volume, CSA and the mean fractional anisotropy of the cervical cord as features that independently influenced the EDSS score using a multivariate regression model. Composite scores, obtained by combining these three features, improved the correlation with clinical scores when compared to the correlations of a single feature. However, DT-MRI is much less commonly acquired than structural MRI. Valsasina et al. [69] explored the regional atrophy of the cervical cord by applying voxel-wise statistics on registered spinal cord segmentations. They used the determined regional atrophy in a multiple regression model, adjusted for age, sex, and cord volume, and showed correlations with clinical scores and patterns of atrophy.

Although a number of composite MRI biomarkers for MS have been proposed, computing morphological features to capture atrophy and combining these features in *linear* and *non-linear regression models* has not been well studied. As well, few works have testing whether combining *multiple spinal cord features* into a single model will provide a better indicator of disability than just using a single feature. Introducing new atrophic features and methods to combine them may assist clinicians in diagnosis, provide insights into disease progression, and serve as a useful composite biomarker.

Few works have specifically examined correlating multiple image features with the progression of MS; however, this concept of using appearance and/or morphological features to predict disability or to detect abnormalities has been successfully demonstrated in other clinical applications. Ward et al. [71] computed 3D shape characteristics from MR images of patients' shoulders and found differences using an ANOVA test. They combined those significant features in a support vector machine (SVM) [23] to distinguish pathology. Using extracted features from DT and dynamic contrast enhanced MRIs, Moradi et al. [53] used a SVM to distinguish between cancerous and normal biopsy samples. Mirzaalian et al. [50] used wavelet transforms extended with medial-based mesh projections extracted from MR images of the thigh muscles and employed a SVM to classify healthy controls from those with chronic obstructive pulmonary disease. Moradi et al. [52] detected prostate cancer

using neural networks and features extracted from the discrete Fourier transform of ultrasound radio-frequency time series data. Tang et al. [65] extracted motion features from ultra sound images to predict patients with or without speech impediments. These works and others suggest that a similar approach may work to predicting disability in MS patients using MRI based features.

We propose novel features extracted from MRI and the corresponding spinal cord segmentation that are potentially more specific to the clinical status than pure area or volume. Using these extracted features, we employ different regression models ranging in complexity and intuitiveness, starting with simple linear regression models, then multiple linear regression models and finally, non-linear non-parametric regression forests. To determine which of our proposed candidate features are useful biomarkers, we explore our data for features that are consistently associated with clinical state. Our results suggest that our proposed features and the more complex regression models are capable of outperforming the predictive abilities of a linear regression model using only spinal cord volume as the explanatory variable.

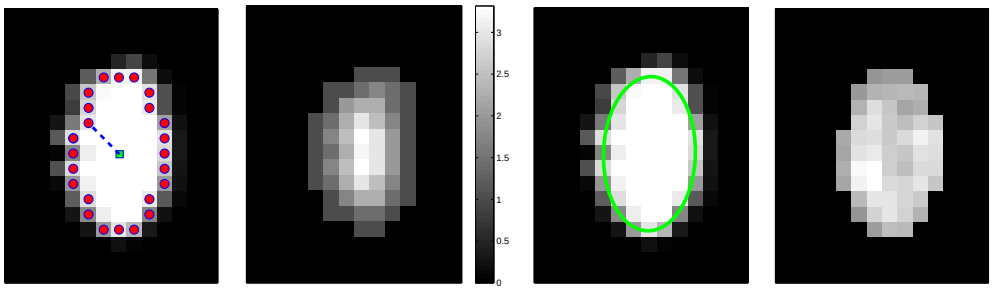
4.2 Methods

In this section we describe our data and the regression problem, examine the new candidate spinal cord features, outline the different types of regression models used, describe our cross-validation set-up, and finally discuss how the clinical scores are computed.

4.2.1 The Data and the Problem

We are given a set of n MRI scans $I = \{I_1, \ldots, I_n\}$ where each 3D MRI scan I_i has a corresponding real number clinical score $y_i \in Y$, and a corresponding spinal cord segmentation $S_i \in S$. The dimensions of I_i and S_i are the same. As described in our previous chapters, each voxel in S_i has a value between 0 and 1, where 0 represents the background and 1 represents the spinal cord. Voxels in S_i that are on the boundary of the spinal cord are assigned a fuzzy value between 0 and 1 that represents an estimated percentage of the voxel that contains spinal cord (i.e., partial volume) [66].

Our objective is to create a model M, using the images I and segmentations S, capable of predicting the patients' clinical scores Y from novel MR images. We extract a set of



(a) (b) (c) (d)

Figure 4.1: Illustrations of the proposed features. (a) The distances (*dashed line*) from the center-of-mass (*center box*) to the boundary voxels (*circles*) make up per_k. (b) The distances to the nearest boundary point from the voxels inside the cord give $dist_k$ (brighter implies farther). (c) An ellipse is fit to the cord. (d) The normalized intensities of the cord are considered in int_k.

features X from I and S that are transformed by model M into values $\hat{Y}$, such that the predicted values $\hat{Y} = M(X)$ estimate the corresponding clinical scores Y.

One approach is to set M as a simple linear regression model with the spinal cord volume as the single explanatory variable X. This is similar to the existing literature where a Pearson's correlation coefficient is computed to measure the linear dependency between the spinal cord volume and clinical score. However, as mentioned in the introduction, this linear dependency using spinal cord volume does not always reveal a strong clinical relationship. We improve on this by deriving new morphological and MRI-based appearance features X and examining ways to combine them in more descriptive models M.

4.2.2 Candidate Features

We describe simple candidate morphological and appearance features X that are potentially sensitive to spinal cord changes. This is not meant to be a comprehensive set of features, but is sufficient to explore the potential of going beyond measuring cord size to predict disability. We first define the commonly used spinal cord volume, which is computed by summing all voxels, including the partial volumes $S_i(m) \in [0, 1]$, in the segmentation, $vol = \sum_{j=1}^{J} S_i(j)$, where J is the total number of voxels in S_i. While spinal cord volume captures a global measure of spinal cord atrophy, we are also interested in features that vary at least partly independently from area or volume, and that are sensitive to spinal cord changes at a local scale.

Our first proposed feature is designed to be more sensitive to local changes in the spinal cord's boundary. On each 2D axial slice of the segmentation S_i, we find voxels on the boundary between the spinal cord and background by considering voxels in S_i with a partial volume greater than 0.5 to be spinal cord. This results in a 2D binary image that we use to extract the cord's boundary voxels. For the k^{th} 2D axial slice of the spinal cord, we take the Euclidean distance between the center-of-mass c_k of the cord's k^{th} cross section, and the spinal cord boundary/perimeter voxels b computed as, $per_k = (d(c_k, b_k^1), \ldots, d(c_k, b_k^{m(k)}))$, where b_k^i represents the i^{th} boundary voxel on the k^{th} slice, and $d(c, b)$ computes the Euclidean distance between the two coordinates (Fig. 4.1a). The number of boundary voxels $m(k)$ can change for each 2D slice. We find the *minimum* distance from the center-of-mass to the boundary voxels in each 2D slice averaged over K 2D slices,

$$per_{\text{min}} = \frac{1}{K} \sum_{k=1}^{K} \min(per_k). \tag{4.1}$$

In a similar way, to compute additional features we replace the "min" function from (4.1) with the mean (per_{mean}), standard deviation (per_{std}), and the max (per_{max}) functions.

We define a related measure that focuses on local changes in 3D by calculating a 3D distance transform from the surface of the segmented spinal cord masked by (or restricted to) the interior region of the cord. To compute the distance transform, we calculate the Euclidean distance between voxels inside the spinal cord and the nearest boundary voxel in 3D. To further differentiate this feature from the *per* features, we consider voxels that contain any partial volume to be spinal cord, which changes the boundary voxels. The distance transform for slice k with q voxels inside the cord is represented as $dist_k = (t_k^1, \ldots, t_k^{q(k)})$ where t_k^i is the distance from the i^{th} voxel inside the cord on the k^{th} slice to the nearest 3D boundary coordinate (Fig. 4.1b). The number of voxels inside the cord, $q(k)$, can change for each 2D slice. In a similar fashion to (4.1), we replace per_k with $dist_k$ and the "min" function with the mean ($dist_{\text{mean}}$), max ($dist_{\text{max}}$), standard deviation ($dist_{\text{std}}$) and the max divided by the mean distance ($dist_{\text{mean}}^{\text{max}}$) function averaged over the K 2D slices. For clarity we formally define,

$$dist_{\text{mean}}^{\text{max}} = \frac{1}{K} \sum_{k=1}^{K} \frac{\max(dist_k)}{\text{mean}(dist_k)}, \tag{4.2}$$

which averages the ratio of the furthest boundary distance by the mean distance.

To compute features that are more robust to local noise, such as small segmentation errors, we fit an ellipse (Fig. 4.1c) to each 2D cross-sectional slice of the segmented spinal cord

and compute the eccentricity (ecc), minor axis (ax_{min}), and major axis (ax_{maj}), averaged over the length the cord.

All the features proposed so far are dependent on the geometrical characteristics of the cord, but we also include features based on the intensities found within the MRI. As the intensity values can vary widely in different MRI scans, we normalize a scan's intensities by its overall 3D scan intensities to produce z-scores. We extract the z-scores of those voxels that are labelled as spinal cord (partial volume > 0.5) and take the mean (int_{mean}) and standard deviation (int_{std}) of the spinal cord intensity values averaged over the K 2D slices (Fig. 4.1d).

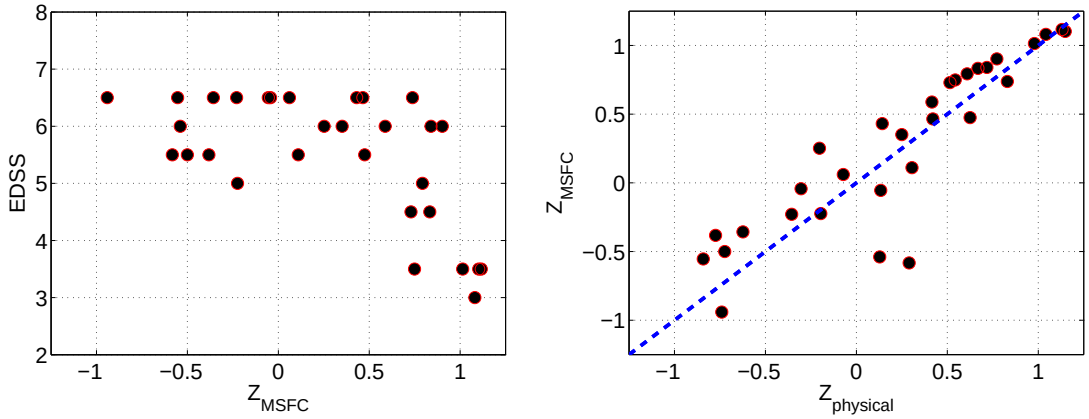
4.2.3 Regression Models

Linear regression [51] employs a linear function to model the relationship between the explanatory variable (e.g. spinal cord volume) and a response variable (clinical score). The parameters of this model are the coefficients β of the explanatory variables and the error term ϵ. These coefficients can be estimated from the data by applying a *least-squares* fitting [73] that minimizes the differences between the response variable and the fitted explanatory variable. A model with only a single explanatory variable x_1, is known as *simple linear regression*, and is one of the simplest models to analyze. Given a dataset with n observations, this produces a straight line, $y_i = \beta_1 x_{i1} + \epsilon_i, i = 1, \ldots, n$. *Multiple linear regression* builds on this by adding r explanatory variables to the model, $y_i = \beta_1 x_{i1} + \cdots + \beta_r x_{ir} + \epsilon_i$.

While these models assume a linearity of the underlying relations, we also explore a more flexible, non-linear, non-parametric model, known as a *regression forest*. A regression forest significantly differs from the previously described models as it is completely learned from the data and makes no assumptions about the underlying distributions [18].

4.2.4 Training and Testing the Models

The models in section 4.2.3, are described in order of increasing complexity. With this added complexity, we increase the potential to accurately model the underlying function, but also increase the difficulty in intuitively understanding the model and increase the likelihood of over-fitting the model to the training data. To reduce the possibility of over-fitting, we divide our data into a training and testing set. Given the relatively small size of our dataset, we use leave-one-out cross-validation. This is repeated for all samples to give us an indication



(a) EDSS vs. Z_{MSFC} (b) Z_{MSFC} vs. Z_{physical}

Figure 4.2: The distributions of scores are shown. (a) The Z_{MSFC} scores have a wider distributions than EDSS scores. As expected, as EDSS decreases, there is a trend for Z_{MSFC} to increase. (b) We remove the cognitive component from Z_{MSFC} to form Z_{physical}, slightly changing the distributions (*deviations from dashed line*).

of the robustness and generalizability of our regression model and chosen features.

4.2.5 Clinical Scores

As discussed in the introduction, the EDSS and the MSFC scores, which we aim to predict from X, are commonly used to quantify clinical disability. We choose to focus on the MSFC score rather than the EDSS score because the MSFC captures disability to which the EDSS score is relatively insensitive, such as arm/hand function. In addition, the EDSS scores tend to exhibit a poor distribution due to the non-linearity of the scale, with many patients clustered between 4.5 and 6.5 (Fig. 4.2a).

The MSFC score tests for: upper extremity function, determined by a 9-hole peg test (9-HPT); walking speed, measured by a timed 25-foot walk (T25W); and cognitive function, evaluated by a paced auditory serial addition test (PASAT). These three tests are shown to vary relatively independently, be sensitive to changes over time, and capture aspects of MS that are not captured in the EDSS score [22]. These components averaged together compose the MSFC score,

$$Z_{\text{MSFC}} = (Z_{\text{9-HPT}} - Z_{\text{T25W}} + Z_{\text{PASAT}})/3 \tag{4.3}$$

where the scores are normalized to produce z-scores using a reference population that includes healthy controls [22].

While this composite score is used to give an overall indication of the progression of multiple sclerosis, we do not expect the cognitive component, Z_{PASAT}, to have a strong causal relation with spinal cord atrophy as the spinal cord is not directly related to cognitive function. We test this by computing the Pearson's correlation coefficient with the cognitive test Z_{PASAT} and spinal cord volume vol, and do not find a significant correlation (r = -0.016, p-value = 0.93). For this reason, we remove Z_{PASAT} and only include the physical disability tests to define a new clinical measurement of physical disability,

$$Z_{\text{physical}} = (Z_{\text{9-HPT}} - Z_{\text{T25W}})/2. \tag{4.4}$$

This combined physical score, Z_{physical}, is the clinical score we use as the response variable for this work. The distribution of values and the changes in correlation between Z_{MSFC} and Z_{physical} are shown in Fig. 4.2b.

4.3 Results

We validate our proposed features and models using 30 3D T1-weighted MRIs acquired with a spoiled gradient echo sequence and an MR field strength of either 1.5 tesla or 3.0 tesla. These scans were gathered from multiple centers and parameters varied by site. Each scan is from a different patient (age ranged from 34 to 64) with secondary progressive MS. This dataset is different than what was used in our previous chapters as the previously used dataset included non-MS patients and the necessary clinical scores were not available.

For each 3D MRI, we have its corresponding clinical score as described in Section 4.2.5 and a segmentation of the spinal cord. To ensure reasonably accurate segmentations, we use the same method to segment the cord as was done in the previous chapters, which we describe in more detail here. This method uses a seeded semi-automatic method similar to Tench et al. [66] where a user-guided region growing algorithm marks the spinal cord voxels with a 1 and the background voxels with a 0. Due to the limited resolution of the MRIs and the small size of the cord, voxels on the boundary of the spinal cord, composed both of spinal cord and background, make up approximately 25% of the total voxels in the cross-sectional area [66]. To give an estimate of the spinal cord area contribution these boundary voxels make, the boundary voxels are assigned a fuzzy value between 0 and 1, computed as

a function of the cord, boundary and cerebrospinal fluid intensities, based on Equation (2) in [66].

The original MRI voxel resolutions were either $0.976 \times 0.976 \times 1$ mm or $0.976 \times 0.976 \times 1.3$ mm, but are normalized via trilinear interpolation to $1 \times 1 \times 1$ mm. When computing our features X, we only consider the first 20 2D slices starting from and including the C3 region and moving inferior, i.e. K=20 in (4.1) and (4.2).

4.3.1 Error Metrics

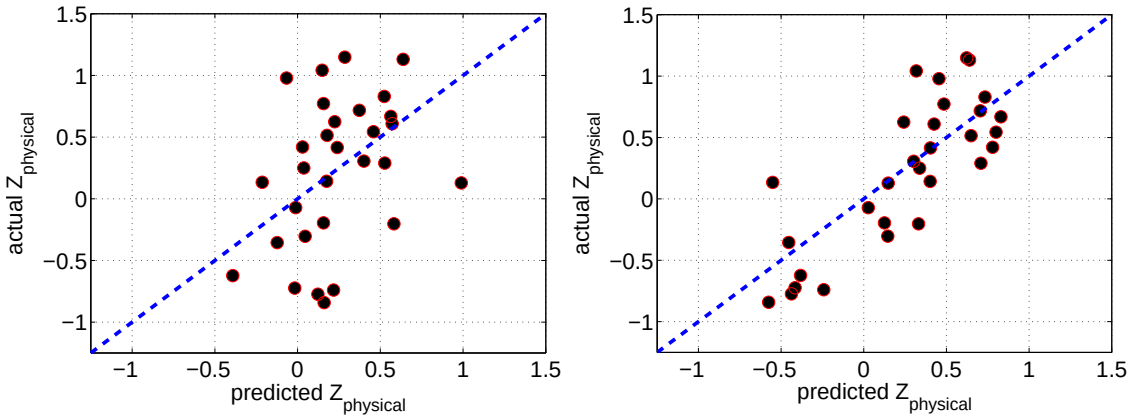
To quantify how closely the predictions $\hat{Y}$ produced by our model are to the true clinical scores Y, we use the following metrics. We compute the *mean absolute error* (MAE) by taking the mean of the absolute difference between the predicted score and the true clinical score, $\text{MAE} = \frac{1}{n}\sum_i^n |\hat{y}_i - y_i|$, giving equal weight to all errors. To get an indication of the variability in the error, we compute the *standard deviation of absolute error* as, $\text{SAE} = \text{std}(|\hat{Y} - Y|)$. To give a higher weight to larger errors, we report the *root mean square error*, $\text{RMSE} = \sqrt{\frac{1}{n}\sum_i^n (\hat{y}_i - y_i)^2}$. MAE, SAE, and RMSE values closer to zero indicate a better model. To indicate the consistency of our predictions, we also compute the *Pearson's correlation coefficient* and its corresponding p-value between the predicted clinical scores $\hat{Y}$ and the true clinical scores Y.

4.3.2 Simple Linear Regression with Spinal Cord Volume

To establish a baseline test on which we aim to improve, we use a simple linear regression model with spinal cord volume as the explanatory variable similar to what is done by Losseff et al. [42]. We compute the volume of the segmented cord (*vol*) and use leave-out-one cross-validation to train our model and test on the omitted volume. As expected from the existing literature [8, 41, 42, 57, 69], we detect a moderate yet statistically significant correlation between volume and clinical score (*vol*: r=0.473, p=0.00824). The predictive ability for a linear regression model using volume as the explanatory variable is reported in Table 4.1 (row 1) and shown in Fig. 4.3a.

4.3.3 Simple Linear Regression with Proposed Features

In our second test, we examine each proposed feature's ability to act as the explanatory variable in a linear model. For each proposed feature in Section 4.2.2, we compute the



(a) Linear regression using cord volume

(b) Regression forest using two features

Figure 4.3: Actual vs. predicted clinical scores are shown. (a) Spinal cord volume *vol* is used as the explanatory variable in a simple linear regression model. (b) A regression forest trained on two selected features, ax_{maj} and per_{min}, demonstrates an improved correlation. Deviations from the *dashed line* are errors.

Pearson's correlation coefficient between the proposed feature and the clinical scores. We find that $ax_{\text{min}}, per_{\text{mean}}, per_{\text{min}}, dist_{\text{max}}^{\text{mean}}$ all provide a slight increase in correlation when compared to *vol*. Of these features, per_{min} shows the strongest improvement in Pearson's correlation (per_{min}: r=0.565, p=0.00115; vs. volume *vol*: r=0.473, p=0.00824) and the p-value of per_{min} survives the Bonferroni correction for multiple testing ($0.00115 < \frac{0.05}{13}$).

We test if per_{min} is a stronger explanatory variable than volume by performing the same cross-validation procedure. We report our results in Table 4.1 (row 2), which demonstrates that not only does per_{min} correlate better than volume, but it gives a more consistent score and is less susceptible to outliers. This is shown by the lower MAE, SAE, and RMSE scores, and higher Pearson's correlation when compared to a model using volume. This suggests that per_{min} may be a better indicator of physical disability than spinal cord volume.

4.3.4 Multiple Linear Regression with Proposed Features

To explore the use of multiple explanatory variables in a linear regression model, using the 13 candidate features described in Section 4.2.2, we form separate models where each feature can either be included or excluded from the model, for a total of $2^{13} = 8192$ possible combinations. To get a sense of which variables generalize well, we test each model using leave-one-out cross-validation. We correct for multiple testing by applying the positive False

Discovery Rate (pFDR) [63] to reduce the likelihood that a positive result is a Type I error. As our goal is to determine if a multiple linear regression model can provide improvements over simple linear regression, we compute how many models result in a RMSE that are less than the RMSE reported using the linear model with the explanatory variable $per_{\min}$ (i.e. RMSE < 0.527). There are 292 such models and from this subset of models, we find the maximum p-value to be 0.00684 with a corresponding q-value of 0.000172. Out of all our tests, there are 749 tests with a p-value less than 0.00684, indicating a low number ($749 \times 0.000172 < 1$) of improved models that are potentially false positives. The features selected from the model with the lowest RMSE are: $best_{7\text{MR}} = \{int_{\text{mn}}, ax_{\min}, per_{\text{mean}}, per_{\max}, per_{\min}, dist_{\max}, dist_{\max}^{\text{mean}}\}$, and the prediction results are reported in Table 4.1 (row 3). We note that this model with multiple features shows a significant reduction in prediction error when compared to the models using a single explanatory variable.

However, as the issue of how best to correct for multiple testing is still an open one, we further examine our models for a more conservative selection of features. We examine what features were consistently selected in the top 25 models. As can been seen in Fig. 4.4, the same five features are selected in nearly every model suggesting these features jointly are useful. Based on this trend, we form a linear regression model using only the consistently selected features, $sel_{5\text{MR}} = \{int_{\text{mean}}, per_{\text{mean}}, per_{\max}, dist_{\max}, dist_{\max}^{\text{mean}}\}$, and report the cross-validated results in Table 4.1 (row 4). While the predictive ability of this model is less than the $best_{7\text{MR}}$ predicting model, this model has two less explanatory variables than the $best_{7\text{MR}}$ model, which may be more generalizable in a novel dataset (even though we cross-validated our dataset). These improvements over the models with a single explanatory variable, suggests that it is useful to combine multiple spinal cord features within a single model.

4.3.5 Non-linear Regression Forest with Proposed Features

In our final tests, we use a non-linear regression forest (RF) implemented with MATLAB's TreeBagger class (R2012a; The MathWorks Inc., Natick, MA). The minimum number of observations per leaf is set to one. All other parameters are left to their default settings except for the number of trees which we describe below. To see if a non-linear model, trained on a single feature can outperform a linear model, we train a RF with 250 trees on each proposed feature from Section 4.2.2. Out of our 13 proposed features, we find that $per_{\min}$ on its own returns superior results when compared to the other models that use only a

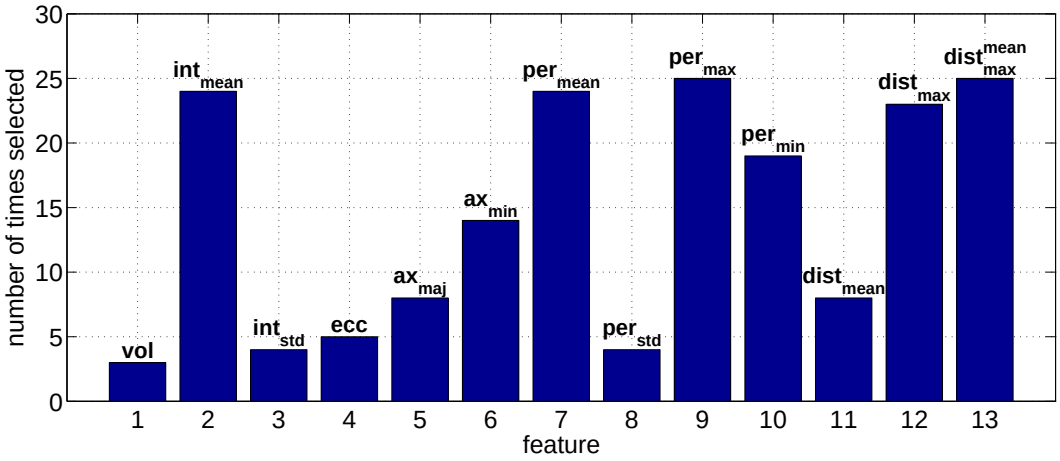


Figure 4.4: The number of times a features was selected in the top (lowest RMSE) 25 multiple linear regression models is graphed. The y-axis shows the number of times the feature was selected and the x-axis is the feature selected. We can see that two features were selected in all the top 25 models, $per_{\max}, dist_{\max}^{\text{mean}}$, two were selected in 24 models, $int_{\text{mean}}, per_{\text{mean}}$, and one was selected in 23 models, $dist_{\max}$. These five features are consistently selected which suggests their general importance in forming the model.

single feature, Table 4.1 (row 5). To consider multiple features in our RF, as was done in Section 4.3.4, we try all possible combinations of features (2^{13}) in a RF. However, to lower computational cost, we use 80 trees with 6-fold (instead of leave-one-out) cross validation when exploring all the feature combinations. We find those features used in the model that produces the lowest RMSE. Correcting for multiple testing using pFRR (sec. 4.3.4), returns less than 1 expected number of false positives.

Similar to Section 4.3.4, we also examine a more conservative selection of features by choosing those features that are consistently in the 25 models with lowest RMSE. We find that the features used in the lowest RMSE model and the features consistently chosen in the 25 lowest RMSE models *are the same*. These selected features are the ax_{maj} (chosen in 24 out of 25 models) and the $per_{\min}$ (chosen in 25 out of 25 models). We train another RF with 250 trees on $sel_{2\text{RF}} = \{ax_{\text{maj}}, per_{\min}\}$ and show leave-one-out cross-validated results that outperform all our previous regression models, reported in Table 4.1 (row 6) and shown in Fig. 4.3b. This demonstrates that select novel morphological features, combined in a non-linear, non-parametric regression model can potentially provide more accurate predictions of MS physical disability than a linear model, and outperforms predictions based on spinal cord volume.

Table 4.1: The *model* column contains the different type of models explored where *linear* represents a linear model, *multiple* represents a multiple linear regression model, and *RF* represents a regression forest model. The *features* column contains the different features the model was trained on, where *vol* represents the volume of the spinal cord, $per_{\min}$ represents the minimal distance to the cord's center-of-mass from the cord's boundary, *best* represents the combination of features that gives the lowest RMSE error, and *sel* are the features consistently selected in our top 25 models. The error metrics we report are the *Mean Absolute Error* (MAE), the *Root Mean Squared Error* (RMSE), the *Standard deviation of Absolute Error* (SAE), the *Pearson's* correlation coefficient r and its corresponding p-value before correction for multiple comparisons.

model	features	MAE	SAE	RMSE	r	p-value
linear	*vol*	0.448	0.326	0.551	0.367	0.0460841
linear	$per_{\min}$	0.444	0.290	0.527	0.464	0.0097723
multiple	$best_{7\text{MR}}$	0.379	0.253	0.453	0.667	0.00005645
multiple	$sel_{5\text{MR}}$	0.414	0.233	0.473	0.617	0.00028511
RF	$per_{\min}$	0.381	0.251	0.453	0.682	0.00003277
RF	$sel_{2\text{RF}}$	**0.293**	**0.201**	**0.353**	**0.803**	0.00000009

4.4 Conclusion

We proposed new morphological and appearance features to capture the subtle changes in a patient's spinal cord as it undergoes atrophy due to multiple sclerosis. These proposed features were combined in a regression model and our results indicate that they are potentially useful imaging biomarkers for multiple sclerosis. When only considering any one particular feature, the distance from the cord's center-of-mass to the cord's boundary, $per_{\min}$, provided the strongest results and was an improvement over spinal cord volume at clinical prediction.

Our results also suggest that combining the selected features in a regression model improves the predictive ability over a simple linear regression model using any one of the tested features, including volume, alone. As well, a non-linear regression forest, trained on select morphological features, appears to be a promising approach to improve on the predictive ability of linear models. To ensure generalizability of our results (i.e. that the proposed biomarkers and models are not specific to our data and that our findings are not due to a Type I error), even though our data came from multiple centers, future work must involve larger datasets representing a greater variety in imaging, pathological, and clinical parameters.

Chapter 5

Conclusions

We have researched and developed two methods to segment the spinal cord from 3D MR images, and explored novel features and regression models to predict the physical disability of MS patients using multiple features extracted from the spinal cord segmentations. In this final chapter, we summarize our contributions, discuss the differences in our segmentation approaches, and finally examine future work.

Our specific contributions are as follows:

- Extending the previous four dimensional search described by Li and Yezzi [40], we proposed a novel spinal cord segmentation algorithm that finds a minimal path in six dimension (three spatial, three shape descriptors) to return a globally optimal spinal cord segmentation. We demonstrated how the principal component analysis shape space can be used to represent arbitrarily high dimensions and addressed some of the non-trivial space and time challenges that occur when moving to higher dimensions. While higher dimensions and/or shape representations can be used, consideration must be given to the high time and memory requirements. Our results demonstrate the feasibility of high-dimensional searches and may prove useful to detect other connected tubular structures.

- We extended Tu's auto-context framework [67] by proposing global geometric features. These features consider the geometric properties and interactions of the candidate shapes created in each iteration of auto-context. This machine learning approach was shown to improve on the existing auto-context framework and outperform our previous spinal cord segmentation method, while requiring less interaction. The concept of

global geometric features may prove to be useful for many other image analysis problems, specifically in those cases where the voxel and approximate size of the object to segment is known, as is commonly the case in medical image analysis problems.

- We presented novel morphological and appearance features to capture the atrophy of the spinal cord and examined combining them into linear and non-linear regression models. Multiple spinal cord-based features were combined into a single model. This improved the model's ability to predict the physical disability of a patient with MS over a model using only the volume of the spinal cord, as is commonly done. In addition, we showed promising improvements to predictive power by using a non-linear regression forest with two of our proposed spinal cord morphological features.

5.1 Comparing the Segmentation Approaches

We compare our two spinal cord segmentation approaches to further understand and clarify the relative advantages and limitations of each approach. Our **first approach** described in Chapter 2, from a computing science perspective, has several desirable properties: a guaranteed global optimal solution, a relatively simple understandable energy function, and an established shape prior extended for probabilistic values. With careful analysis, we can examine this algorithm and propose specific improvements (e.g. a more flexible shape space or additional energy terms) to improve the segmentation. However, balancing between a model that accurately describes the cord and a model that is optimizable becomes a primary concern. To more accurately describe the cord, we can add more edges to the graph and increase our search resolution, but this quickly makes the optimization problematic as it increases both the space and time requirements. But perhaps more importantly, this method (as it is currently formulated - other extensions are possible) has an underlying assumption that we can derive by hand a set of energy terms that will robustly and accurately model the spinal cord. Although learning the appropriate weights and parameters for each energy term may provide improvements [46, 64], the energy terms themselves still are modelled based on the assumptions made by the human designer.

Our second approach to segmentation, described in Chapter 3, is less deterministic and less intuitive than the first: it is difficult to make claims of global optimality and the method to detect and regularize the cord is not easily or intuitively understood. Instead, the algorithm learns the discriminative features and the underlying distributions to both

detect and regularize the spinal cord. The sacrifice of global optimality is supported by the recent works of McIntosh and Hamarneh [47] who emphasized the importance of high fidelity shape spaces and energy functions over maintaining convexity. We can think of our machine learning, auto-context approach as a natural extension to the conclusion of McIntosh [44], who stated that future work on energy functions should focus on "incorporating increasing degrees of learned, contextual knowledge." We moved towards this goal by learning the energy function itself, rather than learning the appropriate weights and parameters for individual hand-crafted energy terms. In addition, this method incorporated a large degree of learned, contextual knowledge through the use of global geometric features.

We examine the common equation in these two approaches for a more direct comparison of where the improvement in Chapter 3 comes from. Both these approaches approximate the probability of a segmentation S given the observed image data I, $p(S|I)$. In Chapter 2 we assumed that the probability of the image $p(I)$ is uniform. Thus, applying Bayes' rule, we can simplify our equation to $p(S|I) \approx p(I|S)p(S)$, where $p(I|S)$ is modelled using a hand-crafted energy function (i.e. gradient and intensity-based features) given in Equation (2.3). The probability of the segmentations $p(S)$ is modelled by the PCA shape space where the allowable shape domain (ASD) is constrained by a hyperrectangle. A probability of zero is assigned for shapes outside of the ASD and uniform probabilities are assigned for shapes inside the ASD. The regularization of the segmentation is taken care of by restrictions to the connectivity of the graph, the connected minimal path and the PCA shape space.

In contrast, the equation in Chapter 3 assumes $p(S|I)$ can be directly learned by introducing an iterative time-step component t such that $p(S^t|I) \approx p(S^t|I, S^{t-1})$. The probability, $p(S^t|I, S^{t-1})$, is non-parametrically learned from the data and both detects and regularizes the cord. The hybrid approach attempted in Chapter 3 demonstrated that the improvements not only come from a better detection of the cord (before regularization), but also from an improved regularizer. Thus, we believe that the superior segmentation accuracy, observed using the augmented auto-context in Chapter 3, comes from removing both the hand-crafted energy terms (Equation 2.3) and the restrictions on the shape space, and replacing them with a learned distribution that jointly detects and regularizes the segmentation.

Essentially, our proposed methods have pushed the boundaries of segmentation algorithms towards approaches that consider more context and improved modelling of the context's distribution. Our results in Chapter 3 suggest that increasing the context and ability to model this context's distribution improves the segmentation results.

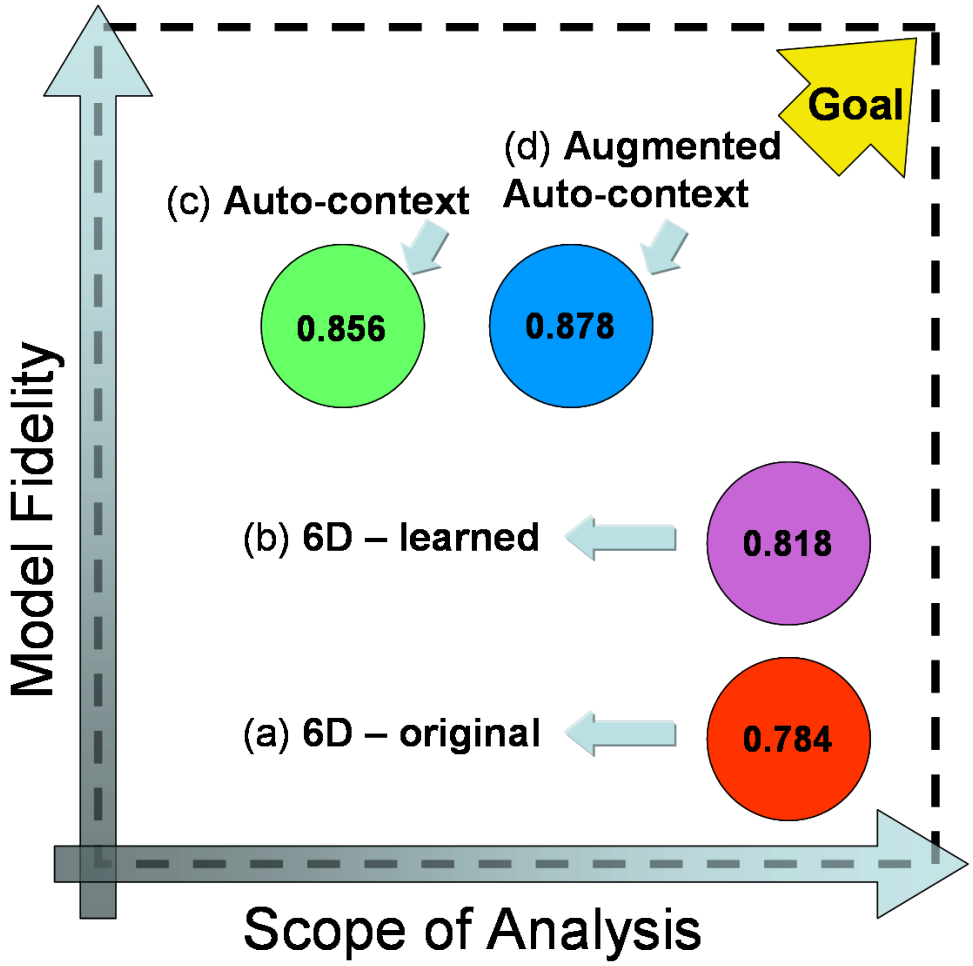


Figure 5.1: An illustration of where the segmentation methods used in our work rank when examining the amount of context (*scope of analysis*) considered and how accurate we model the given scope of analysis (*model fidelity*). The probabilistic Jaccard similarity coefficient (Equation 2.6) between the segmentations produced by the algorithms and the ground truth is measured for 10 MS patients on the same dataset (*results in the circles*). (a) We can see that our 6D method from Chapter 2 has high context, but our assumptions limited the accuracy of the distribution. (b) The accuracy is improved when a trained classifier is used to detect the cord. (c) The auto-context approach of jointly learning the detection and regularization is accurate but only considers the local neighbourhood context. (d) By considering global features, the amount of context is increased yielding the most accurate segmentations.

A visual summary of the segmentation methods used in this work are shown in Fig. 5.1. This figure illustrates where each method stands when considering the amount of context and accuracy of the distributions. In Chapter 2, we obtained a globally optimal solution that considered the entire domain of possible solutions (high *scope of analysis*); however, to make our search optimizable, we made simplifying assumptions in our shape space (three discrete principal component weights) and energy function, resulting in a lower *model fidelity* (Fig. 5.1a). In Chapter 3, we removed the hand-crafted terms in the energy function by using a trained classifier to detect the spinal cord, but the regularization was still done using the PCA shape space and the 6D optimizer. This method has a high *scope of analysis* as it still considers the entire domain of possible solutions, but is limited in modelling an accurate distribution due to the restricted shape space (Fig. 5.1b). The conventional auto-context model proposed by Tu [67] (Chapter 3), learns an accurate distribution to both detect and regularize the cord, but only with a local neighbourhood window and without directly considering the entirety of the segmentation (Fig. 5.1c). By incorporating global features from the segmentations, we can include global context and learn the appearance, context and global-context distributions, which yields the most accurate results (Fig. 5.1d).

Of the two segmentation approaches proposed in Chapter 2 and Chapter 3, we conclude that at least for our data, the approach of learning the underlying rules from the data is superior to my ability to explicitly design such rules. The generalizability of this conclusion is at least partially supported by recent works that rely on learning the underlying rules rather than explicitly encoding them [36,68,75]. As well, these observations agree with recent works that discuss trade-offs between fidelity and optimizabilty of objective functions [25, 47].

5.2 Future Work

We discuss a number of directions for potential future work that could enhance and clarify our conclusions.

- The A* optimizer could greatly benefit from using additional heuristics in the search. For example, we could consider the spatial planes the path must travel through to get to go from the start to the end seed point. The sum of the minimal costs in each plane would be less than or equal to the true minimum cost and thus would be an admissible heuristic. This would likely drastically cut down both the amount of required memory

and the search time, allowing for a larger search domain where higher dimensions and a finer search resolution could be used.

- The types of global geometric features could be expanded for auto-context. We did not explore other types of features as the single feature presented was sufficient for our dataset. More descriptive features may prove to be useful in other types of problems.

- The usefulness of our proposed segmentation methods could be strengthened by comparing how well features extracted from the proposed segmentations indicate or predict physical disability when compared to the features extracted from the ground truth segmentations. If the features extracted from the automated segmentations are stronger indicators than the manual methods, then this would suggest the automated method is more sensitive to changes in the cord and may actually be returning more accurate segmentations.

- Other features could be added to the regression model to predict physical disability. We restricted our search to only intuitive features but other less intuitive features may prove to be better explanatory variables. As well it would be interesting to see if a machine could learn the relevant features that indicate physical disability from the data, and we may end up not relying on the spinal cord segmentations at all.

Bibliography

[1] Xinapse systems. `http://www.xinapse.com/`. Accessed: 2013-07-03.

[2] Shawn Andrews, Chris McIntosh, and Ghassan Hamarneh. Convex multi-region probabilistic segmentation with shape prior in the isometric log-ratio transformation space. *ICCV*, pages 2096–2103, 2011.

[3] Rohit Bakshi, Alan J Thompson, Maria A Rocca, Daniel Pelletier, Vincent Dousset, Frederik Barkhof, Matilde Inglese, Charles RG Guttmann, Mark A Horsfield, and Massimo Filippi. MRI in multiple sclerosis: current status and future prospects. *The Lancet Neurology*, 7(7):615–625, 2008.

[4] Frederik Barkhof, Peter A Calabresi, David H Miller, and Stephen C Reingold. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. *Nature Reviews Neurology*, 5(5):256–266, 2009.

[5] William A Barrett and Eric N Mortensen. Interactive live-wire boundary extraction. *Medical Image Analysis*, 1(4):331–341, 1997.

[6] Peter J Basser and Derek K Jones. Diffusion-tensor MRI: theory, experimental design and data analysis–a technical review. *NMR in Biomedicine*, 15(7-8):456–467, 2002.

[7] Cynthia A Beck, Luanne M Metz, Lawrence W Svenson, and Scott B Patten. Regional variation of multiple sclerosis prevalence in Canada. *Multiple Sclerosis*, 11(5):516–519, 2005.

[8] Beatrice Benedetti, Maria A Rocca, Marco Rovaris, Domenico Caputo, Mauro Zaffaroni, Ruggero Capra, Antonio Bertolotto, Vittorio Martinelli, Giancarlo Comi, and Massimo Filippi. A diffusion tensor MRI study of cervical cord damage in benign and secondary progressive multiple sclerosis patients. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(1):26–30, 2010.

[9] Abi Berger. How does it work?: Magnetic resonance imaging. *BMJ: British Medical Journal*, 324(7328):35, 2002.

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

[11] Carl Bjartmar, Grahame Kidd, Sverre Mörk, Richard Rudick, and Bruce D Trapp. Neurological disability correlates with spinal cord axonal loss and reduced N-acetyl aspartate in chronic multiple sclerosis patients. *Annals of neurology*, 48(6):893–901, 2000.

[12] Neda Changizi and Ghassan Hamarneh. Probabilistic multi-shape representation using an isometric log-ratio mapping. In Tianzi Jiang, Nassir Navab, Josien Pluim, and Max Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2010*, volume 6363 of *Lecture Notes in Computer Science*, pages 563–570. Springer Berlin Heidelberg, 2010.

[13] Min Chen, Aaron Carass, Jennifer Cuzzocreo, P-L Bazin, Daniel S Reich, and Jerry L Prince. Topology preserving automatic segmentation of the spinal cord in magnetic resonance images. *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1737–1740, 2011.

[14] Alastair Compston and Alasdair Coles. Multiple sclerosis. *The Lancet*, 359(9313):1221 – 1231, 2002.

[15] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[16] Kristen Coyne. MRI: A guided tour. National High Magnetic Field Laboratory `http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri/fullarticle.html`. Accessed: 2013-07-03.

[17] Daniel Cremers, Frank R Schmidt, and Frank Barthel. Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–6. IEEE, 2008.

[18] Antonio Criminisi, Jamie Shotton, and Ender Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision*, 7(2-3):81–227, 2011.

[19] Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

[20] Juan José Egozcue, V Pawlowsky-Glahn, G Mateu-Figueras, and C Barceló-Vidal. Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3):279–300, 2003.

[21] Massimo Filippi and Maria A Rocca. Multiple sclerosis: Linking disability and spinal cord imaging outcomes in MS. *Nature Reviews Neurology*, 2013.

[22] JS Fischer, RA Rudick, GR Cutter, and SC Reingold. The multiple sclerosis functional composite measure (MSFC): an integrated approach to MS clinical outcome assessment. *Multiple Sclerosis*, 5(4):244–250, 1999.

[23] Dmitriy Fradkin and Ilya Muchnik. Support vector machines for classification. *Discrete methods in epidemiology*, 70:13–20, 2006.

[24] AlejandroF. Frangi, WiroJ. Niessen, KoenL. Vincken, and MaxA. Viergever. Multiscale vessel enhancement filtering. In WilliamM. Wells, Alan Colchester, and Scott Delp, editors, *Medical Image Computing and Computer-Assisted Interventation MICCAI98*, volume 1496 of *Lecture Notes in Computer Science*, pages 130–137. Springer Berlin Heidelberg, 1998.

[25] Ghassan Hamarneh. The optimizability-fidelty trade-off in image analysis. Technical Report SFU-Summit-10897, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada, September 2011.

[26] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on*, 4(2):100–107, 1968.

[27] William R Hendee and Christopher J Morgan. Magnetic resonance imaging part I-physical principles. *Western Journal of Medicine*, 141(4):491, 1984.

[28] Joseph P. Hornak. The basics of MRI. `http://www.cis.rit.edu/htbooks/mri/`. Accessed: 2013-07-03.

[29] Mark A. Horsfield, Stefania Sala, Mohit Neema, Martina Absinta, Anshika Bakshi, Maria Pia Sormani, Maria Assunta Rocca, Rohit Bakshi, and Massimo Filippi. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: Application in multiple sclerosis. *NeuroImage*, 50(2):446–455, 2010.

[30] Daniel P. Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850–863, 1993.

[31] Paul Jaccard. The distribution of the flora in the alpine zone. *New phytologist*, 11(2):37–50, 1912.

[32] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

[33] Jeremy Kawahara, Chris McIntosh, Roger Tam, and Ghassan Hamarneh. Augmenting auto-context with global geometric features for spinal cord segmentation. In *MICCAI Workshop on Machine Learning in Medical Imaging (MLMI)*, volume 8184 of *Lecture Notes in Computer Science*. Springer, 2013.

[34] Jeremy Kawahara, Chris McIntosh, Roger Tam, and Ghassan Hamarneh. Globally optimal spinal cord segmentation using a minimal path in high dimensions. *International Symposium on Biomedical Imaging (ISBI)*, pages 836–839, 2013.

[35] Jeremy Kawahara, Chris McIntosh, Roger Tam, and Ghassan Hamarneh. Novel morphological and appearance features for predicting physical disability from MR images in multiple sclerosis patients. In *MICCAI Workshop on Computational Methods and Clinical Applications for Spine Imaging*, Lecture Notes in Computer Science. Springer, 2013. Conditionally accepted.

[36] P Kontschieder, P Kohli, J Shotton, and A Criminisi. GeoF: Geodesic forests for learning coupled predictors. *Proc. Computer Vision and Pattern Recognition (CVPR) , IEEE*, 2013.

[37] John F Kurtzke. Rating neurologic impairment in multiple sclerosis an expanded disability status scale (EDSS). *Neurology*, 33(11):1444–1444, 1983.

[38] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[39] Max WK Law, Gregory J Garvin, Sudhakar Tummala, KengYeow Tay, Andrew E Leung, and Shuo Li. Gradient competition anisotropy for centerline extraction and segmentation of spinal cords. In *Information Processing in Medical Imaging*, pages 49–61. Springer, 2013.

[40] H Li and A Yezzi. Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines. *IEEE TMI*, 26(9):1213–1223, 2007.

[41] X Lin, CR Tench, B Turner, LD Blumhardt, and CS Constantinescu. Spinal cord atrophy and disability in multiple sclerosis over four years: application of a reproducible automated technique in monitoring disease progression in a cohort of the interferon β-1a (Rebif) treatment trial. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(8):1090–1094, 2003.

[42] NA Losseff, SL Webb, JI O'riordan, R Page, L Wang, GJ Barker, PS Tofts, WI McDonald, DH Miller, and AJ Thompson. Spinal cord atrophy and disability in multiple sclerosis a new reproducible and sensitive MRI method with potential to monitor disease progression. *Brain*, 119(3):701–708, 1996.

[43] Ranbir S Mann, Cris S Constantinescu, and Christopher R Tench. Upper cervical spinal cord cross-sectional area in relapsing remitting multiple sclerosis: Application of a new technique for measuring cross-sectional area on magnetic resonance images. *Journal of Magnetic Resonance Imaging*, 26(1):61–65, 2007.

[44] Chris McIntosh. *Energy functionals for medical image segmentation: choices and consequences*. PhD thesis, Simon Fraser University, 2011.

[45] Chris McIntosh and Ghassan Hamarneh. Spinal crawlers: deformable organisms for spinal cord segmentation and analysis. In *Proceedings of the 9th international conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I*, MICCAI'06, pages 808–815, Berlin, Heidelberg, 2006. Springer-Verlag.

[46] Chris McIntosh and Ghassan Hamarneh. Is a single energy functional sufficient? Adaptive energy functionals and automatic initialization. In Nicholas Ayache, Sbastien Ourselin, and Anthony Maeder, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2007*, volume 4792 of *Lecture Notes in Computer Science*, pages 503–510. Springer Berlin Heidelberg, 2007.

[47] Chris McIntosh and Ghassan Hamarneh. Medial-based deformable models in nonconvex shape-spaces for medical image segmentation using genetic algorithms. *IEEE Transactions on Medical Imaging (IEEE TMI)*, 31(1):33–50, 2012.

[48] Chris McIntosh, Ghassan Hamarneh, Matthew Toom, and Roger C. Tam. Spinal cord segmentation for volume estimation in healthy and multiple sclerosis subjects using crawlers and minimal paths. In *Proceedings of the 2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*, HISB '11, pages 25–31, Washington, DC, USA, 2011. IEEE Computer Society.

[49] David H Miller, Frederik Barkhof, Joseph A Frank, Geoffrey JM Parker, and Alan J Thompson. Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. *Brain*, 125(8):1676–1695, 2002.

[50] Hengameh Mirzaalian, Ghassan Hamarneh, Bahareh HajGhanbari, and W. Darlene Reid. 3D shape analysis of the knee extensor and flexor muscles in patients with COPD using mesh projection-based features. In *Medical Image Computing and Computer-Assisted Intervention Workshop on Mesh Processing in Medical Image Analysis (MICCAI MeshMed)*, pages 1–8, 2011.

[51] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*, volume 821. Wiley, 2012.

[52] M. Moradi, P. Mousavi, D.R. Siemens, E.E. Sauerbrei, P. Isotalo, A. Boag, and P. Abolmaesumi. Discrete fourier analysis of ultrasound RF time series for detection of prostate cancer. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 1339–1342, 2007.

[53] Mehdi Moradi, Septimiu E. Salcudean, Silvia D. Chang, Edward C. Jones, Nicholas Buchan, Rowan G. Casey, S. Larry Goldenberg, and Piotr Kozlowski. Multiparametric MRI maps for detection and grading of dominant prostate tumors. *Journal of Magnetic Resonance Imaging*, 35(6):1403–1413, 2012.

[54] Jin Nakahara, Michiko Maeda, Sadakazu Aiso, and Norihiro Suzuki. Current concepts in multiple sclerosis: autoimmunity versus oligodendrogliopathy. *Clinical reviews in allergy & immunology*, 42(1):26–34, 2012.

[55] GJ Nijeholt, MA Van Walderveen, JA Castelijns, JH Van Waesberghe, CH Polman, P Scheltens, PF Rosier, PJ Jongen, and F Barkhof. Brain and spinal cord abnormalities in multiple sclerosis. correlation between MRI parameters, clinical subtypes and symptoms. *Brain*, 121(4):687–697, 1998.

[56] M Poon, G Hamarneh, and R Abugharbieh. Live-vessel: extending livewire for simultaneous extraction of optimal medial and boundary paths in vascular images. *MICCAI*, pages 444–451, 2007.

[57] MA Rocca, MA Horsfield, S Sala, M Copetti, P Valsasina, S Mesaros, V Martinelli, D Caputo, T Stosic-Opincal, J Drulovic, G Comi, and M Filippi. A multicenter assessment of cervical cord atrophy among MS clinical phenotypes. *Neurology*, 76(24):2096–2102, 2011.

[58] Fritz Schick. Whole-body MRI at high field: technical limits and clinical potential. *European radiology*, 15(5):946–959, 2005.

[59] B D Schmit and M K Cole. Quantification of morphological changes in the spinal cord in chronic spinal cord injury using magnetic resonance imaging. *Conf Proc IEEE Eng Med Biol Soc*, pages 4425–8, 2004.

[60] Julia A Schnabel, Liqun Wang, and Simon R Arridge. Shape description of spinal cord atrophy in patients with MS. *Comput Assist Radiol ICS*, 1124:286–291, 1996.

[61] VL Stevenson, SM Leary, NA Losseff, GJM Parker, GJ Barker, Y Husmani, DH Miller, and AJ Thompson. Spinal cord atrophy and disability in MS a longitudinal study. *Neurology*, 51(1):234–238, 1998.

[62] VL Stevenson, DH Miller, SM Leary, M Rovaris, F Barkhof, B Brochet, V Dousset, M Filippi, R Hintzen, X Montalban, et al. One year follow up study of primary and transitional progressive multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 68(6):713–718, 2000.

[63] John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

[64] Martin Szummer, Pushmeet Kohli, and Derek Hoiem. Learning CRFs using graph cuts. *ECCV 2008*, pages 582–595, 2008.

[65] Lisa Tang, Ghassan Hamarneh, and Tim Bressmann. A machine learning approach to tongue motion analysis in 2D ultrasound image sequences. In *Machine Learning in Medical Imaging*, pages 151–158. Springer, 2011.

[66] Christopher R Tench, Paul S Morgan, and Cris S Constantinescu. Measurement of cervical spinal cord cross-sectional area by MRI using edge detection and partial volume correction. *Journal of Magnetic Resonance Imaging*, 21(3):197–203, 2005.

[67] Zhuowen Tu. Auto-context and its application to high-level vision tasks. In *Computer Vision and Pattern Recognition. CVPR. IEEE Conference on*, pages 1–8. IEEE, 2008.

[68] Zhuowen Tu and Xiang Bai. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(10):1744–1757, 2010.

[69] Paola Valsasina, Maria A Rocca, Mark A Horsfield, Martina Absinta, Roberta Messina, Domenico Caputo, Giancarlo Comi, and Massimo Filippi. Regional cervical cord atrophy and disability in multiple sclerosis: A voxel-based analysis. *Radiology*, 266(3):853–861, 2013.

[70] Shijun Wang and Ronald M Summers. Machine learning and radiology. *Medical image analysis*, 2012.

[71] Aaron D Ward, Ghassan Hamarneh, Reem Ashry, and Mark E Schweitzer. 3D shape analysis of the supraspinatus muscle: A clinical study of the relationship between shape and pathology. *Academic radiology*, 14(10):1229–1241, 2007.

[72] Eric W. Weisstein. Correlation coefficient. MathWorld–A Wolfram Web Resource `http://mathworld.wolfram.com/CorrelationCoefficient.html`. Accessed: 2013-07-03.

[73] Eric W. Weisstein. Least squares fitting. MathWorld–A Wolfram Web Resource `http://mathworld.wolfram.com/LeastSquaresFitting.html`. Accessed: 2013-07-03.

[74] L Weizman, M Freiman, and L Joskowicz. Implementation of weighted dijkstras shortest-path algorithm for n-D images. *The Insight Journal*, 2009.

[75] Yefeng Zheng, Maciej Loziczonek, Bogdan Georgescu, S Kevin Zhou, Fernando Vega-Higuera, and Dorin Comaniciu. Machine learning based vesselness measurement for coronary artery segmentation in cardiac CT volumes. *SPIE Medical Imaging*, page 79621K, 2011.

[76] Robert Zivadinov and Rohit Bakshi. Central nervous system atrophy and clinical status in multiple sclerosis. *Journal of Neuroimaging*, 14(s3):27S–35S, 2004.