Summer 9-1-2013

Fast and Robust Automatic Segmentation Methods for MR Images of Injured and Cancerous Tissues

Vanessa Karlen Tidwell
Washington University in St. Louis

Washington University in St. Louis

School of Engineering and Applied Science

Department of Electrical and Systems Engineering

Dissertation Examination Committee:
Arye Nehorai, Chair
Martin Arthur
Joel Garbow
Jr-Shin Li
Hiro Mukai
Sheng-Kwei Song

Fast and Robust Automatic Segmentation Methods for MR Images of Injured and Cancerous Tissues

by

Vanessa Karlen Tidwell

A dissertation presented to the Graduate School of Arts and Sciences
of Washington University in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

August 2013
St. Louis, Missouri

Contents

List of Figures

List of Tables

Acknowledgments

I can never hope to name here all of the people who have aided me and supported me in the quest to earn my doctorate.

I would like to start by thanking my advisor, Arye Nehorai, for his vision and guidance. The work presented herein is the fruit of the collaborative environment he has fostered, not just among his students, but also among researchers from varied disciplines and universities.

My research has been driven by interdisciplinary collaboration. The medical researchers who have worked with me, including Joel Garbow, Sasha Krupnick, Joong Kim, 'Victor' Song, and David Gutman, have been invaluable, and they have shaped my research in unexpected ways.

The tools that my professors provided through my course work laid the groundwork for all of my research. My first exposure to image analysis was through a class taught by Martin Arthur. Jr-Shin Li first introduced me to optimization.

I am grateful to all the members of my dissertation committee for the time and effort they have given. Their expertise, support, and feedback has greatly impacted my research and this dissertation.

Throughout my graduate studies, my labmates and fellow students have been indispensable. I am incredibly grateful for the support I have received from each of them. I would particularly like to thank Patricio La Rosa and Pinaki Sarder for their mentorship.

The support of my family and friends has been instrumental in all my successes. My parents started me on this path by always emphasizing the importance of education, and by making sacrifices to ensure that I received a good one. I am indebted to my entire family, most especially both my parents and my parents-in-law, who have been incredibly supportive during this journey. I am thankful also for all of the friends who have encouraged me through this long process, particularly Mara Delcamp.

I have to thank my son, Ethan, who almost derailed my graduation, but who has also been my greatest inspiration for making it to this point.

Finally, and most crucially, I have to thank my husband, Terry. I can't imagine enduring the trials and tribulations of graduate school without his support, encouragement, love, and faith.

Thank you, all!

Vanessa Karlen Tidwell

Washington University in Saint Louis
August 2013

This work was made possible by support from a National Science Foundation (NSF) Graduate Research Fellowship, a Mr. and Mrs. Spencer T. Olin Fellowship for Women in Graduate Study; NSF Grant CCF-0963742; the NIH/NCI Small Animal Imaging Resource Program (U24 CA83060); the Alvin J. Siteman Cancer Center at Washington University in St. Louis, an NCI Comprehensive Cancer Center (P30 CA91842); an NIH/NCI Grant (KO8 CA131097); an American Thoracic Society/Lungevity Foundation Research Grant; and an NIH Grant (NS047592).

Dedicated to Terry.

ABSTRACT OF THE DISSERTATION

Fast and Robust Automatic Segmentation Methods for MR Images of Injured and Cancerous Tissues

by

Vanessa Karlen Tidwell

Doctor of Philosophy in Electrical Engineering

Washington University in St. Louis, August 2013

Professor Arye Nehorai, Chair

Magnetic Resonance Imaging (MRI) is a key medical imaging technology. Through *in vivo* soft tissue imaging, MRI allows clinicians and researchers to make diagnoses and evaluations that were previously possible only through biopsy or autopsy. However, analysis of MR images by domain experts can be time-consuming, complex, and subject to bias. The development of automatic segmentation techniques that make use of robust statistical methods allows for fast and unbiased analysis of MR images.

In this dissertation, I propose segmentation methods that fall into two classes—(a) segmentation via optimization of a parametric boundary, and (b) segmentation via multistep, spatially constrained intensity classification. These two approaches are applicable in different segmentation scenarios. Parametric boundary segmentation is useful and necessary for segmentation of noisy images where the tissue of interest has predictable shape but poor boundary delineation, as in the case of lung with heavy or diffuse tumor. Spatially constrained intensity

classification is appropriate for segmentation of noisy images with moderate contrast between tissue regions, where the areas of interest have unpredictable shapes, as is the case in spinal injury and brain tumor. The proposed automated segmentation techniques address the need for MR image analysis in three specific applications: (1) preclinical rodent studies of primary and metastatic lung cancer (approach (a)), (2) preclinical rodent studies of spinal cord lesion (approach (b)), and (3) postclinical analysis of human brain cancer (approach (b)).

In preclinical rodent studies of primary and metastatic lung cancer, respiratory-gated MRI is used to quantitatively measure lung-tumor burden and monitor the time-course progression of individual tumors. I validate a method for measuring tumor burden based upon average lung-image intensity. The method requires accurate lung segmentation; toward this end, I propose an automated lung segmentation method that works for varying tumor burden levels. The method includes development of a novel, two-dimensional parametric model of the mouse lungs and a multifaceted cost function to optimally fit the model parameters to each image. Results demonstrate a strong correlation (0.93), comparable with that of fully manual expert segmentation, between the automated method's tumor-burden metric and the tumor burden measured by lung weight.

In preclinical rodent studies of spinal cord lesion, MRI is used to quantify tissues in control and injured mouse spinal cords. For this application, I propose a novel, multistep, multidimensional approach, utilizing the Classification Expectation Maximization (CEM) algorithm, for automatic segmentation of spinal cord tissues. In contrast to previous methods, my proposed method incorporates prior knowledge of cord geometry and the distinct information contained in the different MR images gathered. Unlike previous approaches, the algorithm is shown to remain accurate for whole spinal cord, white matter, and hemorrhage

segmentation, even in the presence of significant injury. The results of the method are shown to be on par with expert manual segmentation.

In postclinical analysis of human brain cancer, access to large collections of MRI data enables scientifically rigorous study of cancers like glioblastoma multiforme, the most common form of malignant primary brain tumor. For this application, I propose an efficient and effective automated segmentation method, the Enhanced Classification Expectation Maximization (ECEM) algorithm. The ECEM algorithm is novel in that it introduces spatial information directly into the classical CEM algorithm, which is otherwise spatially unaware, with low additional computational complexity. I compare the ECEM's performance on simulated data to the standard finite Gaussian mixture EM algorithm, which is not spatially aware, and to the hidden-Markov random field EM algorithm, a commonly-used spatially aware automated segmentation method for MR brain images. I also show sample results demonstrating the ECEM algorithm's ability to segment MR images of glioblastoma.

Chapter 1

Introduction

> "Without the aid of statistics nothing like real medicine is possible."
> (Pierre Charles Alexandre Louis, 1837 [38])

Magnetic Resonance Imaging (MRI) is a key medical imaging technology. Through *in vivo* soft tissue imaging, MRI allows clinicians and researchers to make diagnoses and evaluations that were previously possible only through biopsy or autopsy. Moreover, the high-quality scans produced by MRI have the potential to provide crucial new insights into complex biological processes. Therefore, a critical research challenge faced in a wide number of medical applications is to correctly and reliably interpret MR image data. By addressing this challenge via precise characterization and quantification of soft tissue, clinicians and researchers will be able to make decisions based on statistically sound measurement, rather than subjective interpretation.

1.1 MR Imaging

MRI has gained widespread use in medicine because of its ability to image tissue without radiation, unlike imaging technologies like computed tomography (CT) or positron emission tomography (PET). Instead, MRI works by measuring the response of tissue to radio frequency (RF) pulse sequences in the presence of a strong, uniform magnetic field. Its importance was underscored by the 2003 Nobel Prize in Physiology and Medicine awarded to Paul Lauterbur and Sir Peter Mansfield for their pioneering work in the field of MRI.

An MRI scan typically produces a series of 2D image 'slices,' which together form a 3D image. Each 2D slice has depth in addition to width and height; therefore, each 2D image is composed of voxels that correspond to 3D volumes within the imaged region. The image intensity at each voxel gives the magnitude of the corresponding tissue volume's response to the RF pulse sequence. By varying the pulse sequence or adding contrast agents, different MR images can be acquired. For instance, two basic MRI scans are T_1- and T_2-weighted scans, which record the longitudinal and transverse relaxation times of tissue voxels, respectively. Numerous other specialized MRI scans exist.

The various MRI scanning methods can be used to differentiate tissues based on their structural and molecular properties. For instance, diffusion weighted imaging (DWI) captures information about the diffusion characteristics of tissues—e.g., white matter is highly anisotropic in its diffusivity, while gray matter is more or less isotropic, which translates to a difference in DW image intensity between the two tissue types. Thus, in some directions, white matter is bright, in others dark, while gray matter has a medium intensity in all directions. The various MRI scans can each differentiate with greater or lesser contrast between different tissues. Frequently, several MRI scans are used during a single scanning session to produce multiple 'channels' of data, all corresponding to the same imaged region.

MRI technology continues to develop, as does its range of medical applications. Advances in MR imaging include the development of new contrast agents that selectively target specific tissue types, new pulse sequences that better differentiate tissues of interest, and the use of stronger magnetic fields to increase signal to noise ration and image resolution. These factors have given clinicians and researchers an unparalleled wealth of data. However, the sheer quantity and complexity of MR data makes interpretation difficult.

Analysis of MR images by domain experts can be time-consuming, complex, and subject to bias. A common first step for analysis is the segmentation of regions of interest for quantification and characterization. However, the information necessary to segment a region of interest may be masked by noise or spread among several different channels. Multiple channels present an especially difficult challenge for a manual segmenter. It is difficult for a manual segmenter to jointly consider even two different channels of data, let alone data that may contain numerous channels. In an attempt to simplify the data for manual segmentation, various computed maps have been developed that aggregate multi-channel data into a single

image. However, by its very nature this transformation is lossy, i.e., information is discarded. The development of automatic segmentation techniques that make use of robust statistical methods is necessary, not only to save the time and effort currently needed for manual segmentation, but also to overcome the problems of noisy, multichannel data and allow for the fast and unbiased analysis of MR images.

1.2 Contributions

In this dissertation, I propose segmentation methods that fall into two classes—(a) segmentation via optimization of a parametric boundary, and (b) segmentation via multistep, spatially constrained intensity classification. These two approaches are applicable in different segmentation scenarios. Parametric boundary segmentation is useful and necessary for segmentation of noisy images where the tissue of interest has predictable shape but poor boundary delineation, as in the case of lung with heavy or diffuse tumor. Spatially constrained intensity classification is appropriate for segmentation of noisy images with moderate contrast between tissue regions, where the areas of interest have unpredictable shapes, as is the case in spinal injury and brain tumor. The proposed automated segmentation techniques address the need for MR image analysis in three specific applications: (1) preclinical rodent studies of primary and metastatic lung cancer (approach (a)), (2) preclinical rodent studies of spinal cord lesion (approach (b)), and (3) postclinical analysis of human brain cancer (approach (b)).

In preclinical rodent studies of primary and metastatic lung cancer, respiratory-gated MRI is used to quantitatively measure lung-tumor burden and monitor the time-course progression of individual tumors. In Chapter 2, I validate a method for measuring tumor burden based upon average lung-image intensity. The method requires accurate lung segmentation; toward this end, I propose an automated method that accurately segments lungs with varying tumor burden levels. This proposed segmentation method includes development of a novel, two-dimensional parametric model of the mouse lungs and a multifaceted cost function to optimally fit the model parameters to each image. This chapter was previously published in a peer-reviewed journal—minor changes have been made to the published version.

In preclinical rodent studies of spinal cord lesion, MRI is used to quantify tissues in control and injured mouse spinal cords. In Chapter 3, I propose a novel, multistep, multidimensional approach, utilizing the Classification Expectation Maximization (CEM) algorithm, for automatic segmentation of spinal cord tissues. While methods have previously been proposed for whole spinal cord and white matter segmentation of uninjured cords, the proposed algorithm quickly and accurately generates whole spinal cord, white matter, and also hemorrhage segmentations, even in the presence of significant injury. In contrast to previous methods, my proposed method incorporates prior knowledge of cord geometry and the distinct information contained in the different MR images gathered. This chapter was previously published in a peer-reviewed journal—no substantive changes have been made from the published version.

In postclinical analysis of human brain cancer, access to large collections of MRI data enables scientifically rigorous study of cancers like glioblastoma multiforme, the most common form of malignant primary brain tumor. In Chapter 4, I expand upon the method in Chapter 3 with the development of the Enhanced CEM (ECEM) algorithm. My proposed ECEM algorithm incorporates spatial information directly into the CEM framework, which is otherwise spatially unaware, with minimal additional computational complexity, allowing efficient and effective automated segmentation of large data collections.

Chapter 2

Quantitative Analysis of Tumor Burden in Mouse Lung via MRI

This chapter was previously published in *Magnetic Resonance in Medicine*:
V. K. Tidwell, J. R. Garbow, A. S. Krupnick, J. A. Engelbach, and A. Nehorai. Quantitative analysis of tumor burden in mouse lung via MRI. *Magnetic Resonance in Medicine*, 67:572–579, 2012.
Minor changes have been made since publication, including the addition of Section 2.4.

The proposed methodology was also published in *Nature Protocols*:
A. S. Krupnick / V. K. Tidwell, J. A. Engelbach, V. V. Alli, A. Nehorai, M. You, H. G. Vikis, A. E. Gelman, D. Kreisel, J. R. Garbow. Quantitative monitoring of murine lung tumors by magnetic resonance imaging. *Nature Protocols*, 7:128–142, 2012.

I designed, implemented, and tested the parametric lung model and the automatic segmentation algorithm. A.S.K., J.R.G., and I designed the imaging experiments, performed the data analysis, and wrote the manuscript; J.A.E. collected all the MR images; A.E.G. and D.K. helped with experimental design; A.N. helped with data analysis; V.V.A. provided technical support; M.Y. and H.G.V. provided lung tumor-bearing animals and were instrumental in initial efforts to use MRI for monitoring primary mouse lung tumors.

Lung cancer is the leading cause of cancer death in the United States. Despite recent advances in screening protocols, the majority of patients still present with advanced or disseminated disease. Preclinical rodent models provide a unique opportunity to test novel therapeutic drugs for targeting lung cancer. Respiratory-gated MRI is a key tool for quantitatively measuring lung-tumor burden and monitoring the time-course progression of individual tumors in mouse models of primary and metastatic lung cancer. However, quantitative analysis of lung-tumor burden in mice by MRI presents significant challenges. Herein, a method for measuring tumor burden based upon average lung-image intensity is described and validated. The method requires accurate lung segmentation; its efficiency and throughput would be greatly aided by the ability to automatically segment the lungs. A technique for automated lung segmentation in the presence of varying tumor burden levels is presented. The method includes development of a new, two-dimensional parametric model of the mouse lungs and a multi-faceted cost function to optimally fit the model parameters to each image. Results demonstrate a strong correlation (0.93), comparable with that of fully manual expert segmentation, between the automated method's tumor-burden metric and the tumor burden measured by lung weight.

2.1 Introduction

Lung cancer is the leading cause of cancer death in the United States [45]. In 2006, the most recent year for which statistics are available, nearly 200,000 men and women were diagnosed with lung cancer, and almost 160,000 people died from the disease [59]. Despite recent advances in screening protocols, the majority of patients still present with advanced or disseminated disease [57]. While early detection might offer the potential to improve patient survival, the lack of adequate adjuvant therapy after surgical resection hampers long-term survival [25]. Preclinical rodent models provide a unique opportunity to test novel therapeutic drugs to target lung cancer.

The ability to noninvasively record patterns of lung-tumor growth and response to therapy *in situ*, rather than in orthotopically implanted flank tumors, would greatly enhance the utility of small-animal models of lung pathology. This is especially true in light of recent demonstrations that subcutaneous malignancies may undergo progressive growth and regression after

the development of an anti-tumor immune response [6]. A major limitation in the study of murine thoracic tumors, and a potential reason for the paucity of such studies, is the difficulty in the detection and serial growth analysis of malignant and premalignant lung lesions. Unlike injected flank tumors or primary subcutaneous malignancies, which can be palpated and whose growth can be measured with calipers, the ability to monitor tumor growth or response to therapy in lung is limited [12, 7, 42, 22, 18, 63, 50, 23]. Serial measurement of lung tumors requires *in vivo* imaging. While high-resolution microCT is a valuable imaging modality for studying murine lung [33], the scan itself delivers a significant dose of radiation, which can markedly affect tumor growth and tumor immune response. In many studies, small-animal MRI, which employs only non-ionizing radiation, is the imaging modality of choice for characterizing lung-tumor growth and therapeutic response [20]. Recently, we have demonstrated the use of respiratory-gated MRI to quantitatively measure lung-tumor burden and to monitor the time-course progression of individual tumors in mouse models of primary and metastatic lung cancer [7, 22, 23].

Analysis of tumor burden, particularly for heavy or diffuse tumor, by MRI presents significant challenges beyond those associated with data collection. In our previous studies [7, 22, 23], we visually identified individual tumors or groups of tumors (bright signal against the background of dark lung), encircled these tumors with appropriate regions of interest, and measured the corresponding volumes of the identified regions. While time consuming, this approach works well for well-defined tumor masses (Fig. 2.1.b) and the volumes so-derived correlate well with tumor volumes measured histologically. However, this type of process is impractical for diffuse metastatic disease that results in the replacement of the majority of lung parenchyma with tumor (Fig. 2.1.c). Instead, taking advantage of the large difference in MR image intensity between tumor and healthy lung parenchyma, we propose average lung-image intensity as a quantitative measure of tumor burden. (A related metric, the hyperintense-to-total lung volume (HTLV) ratio, has been used to quantify inflammation in an inflammation-mediated lung injury mouse model [54]). Herein, we describe the implementation and validation of such an approach, in which tumor burden, derived from MR lung-image intensity, is correlated with lung mass, which has recently been used as a quantitative measure of tumor in mice [19].

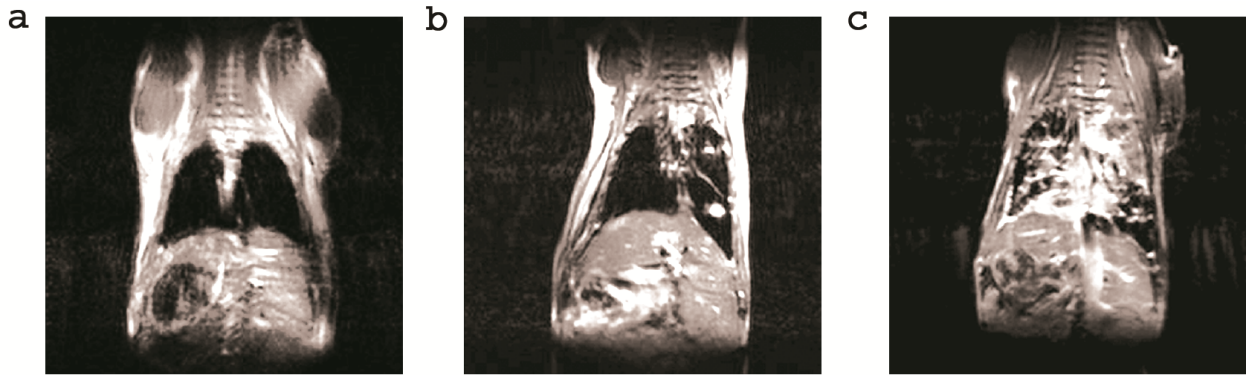

Figure 2.1: Example MRI slices for (a) control mouse with no visible lung tumor, (b) mouse with several discrete lung tumors, and (c) mouse with diffuse metastatic tumor.

A key to the success of our approach for measuring tumor burden is the ability to accurately and reproducibly segment the lungs across the many slices of a 2D multi-slice image. In our 0.5 mm-thick, coronal-slice images, lungs are often represented in 15-20 total slices. As with drawing regions of interest around individual tumors, the manual segmentation of lungs can be slow and time-consuming. The efficiency and throughput of the analysis would be greatly aided by the ability to automatically segment the lungs.

A variety of algorithms for automated and semi-automated tissue segmentation have previously been developed for and applied to lung MR images [49, 44, 36, 37, 4, 53], though none have been applied to the segmentation of lung in the presence of either heavy tumor burden or diffuse tumor. These methods generally rely on the high contrast between healthy lung tissue, which has very low intensity in MR images, and surrounding tissue. Due to the strong intensity gradients at the lung boundary, active contours have been applied successfully in healthy lungs [49, 44]. Threshold-based methods have also been developed [53]. However, these methods are not appropriate for segmentation of lungs with diffuse tumor (Fig. 2.1.c), as the intensity characteristics upon which they rely may not be valid in such images. For example, lung edges may be weak or undetectable, as in the upper-right quadrant of the lung in Figure 2.1.c. Motion artifacts and partial volume effects can lead to elevated intensity levels in voxels within the lungs, contributing an additional source of potential error for threshold-based methods.

Model-based lung segmentation methods have also been proposed [36, 37]. However, the method proposed is suitable only for coarse segmentation of a collection of objects, rather than locally accurate segmentation of a single object, as in our study. Finally, atlas-based 3D segmentation methods have been developed [31, 35]; however, such methods assume that the image area to be segmented is very similar to the atlas, and can be aligned with the atlas through a series of registration steps, which may not be the case for tumor-filled lung. For lungs with diffuse tumor (Fig. 2.1.c), new segmentation methods are required.

To address the challenges of lung segmentation in the presence of varying tumor burden, we developed a new, two-dimensional parametric model of the mouse lungs. The model preserves the overall shape of the lungs, avoiding the inclusion or exclusion of large sections of lung that might occur with non-parametric, threshold-based approaches or edge-detection methods. The parameters of this model are iteratively fit to each MRI slice by utilizing optimization of a multi-faceted cost function. This cost function is novel in that it is evaluated as a function of the intensity distributions both inside and outside the parametric model. While specifically developed and tested in mouse lung, we expect that this new algorithm will have broad application to a variety of segmentation problems.

The dual goals of this chapter are to: 1) validate average lung-image intensity as a quantitative measure of lung-tumor burden and 2) develop and validate a new algorithm for automated lung segmentation. MRI measurements of lung-tumor burden are validated by correlating lung-image intensities with corresponding lung weights, while the results of automated segmentation are validated by direct comparison with manual image segmentations performed by a series of four experts. Excellent congruence is observed between lung volumes derived from the automated and manual lung segmentations, and average image intensities derived from these segmentations correlate well with measured lung weights.

In Section 2.2, we introduce the proposed parametric lung model and segmentation algorithm. In Section 2.3, we present both manual and automatic MRI-based tumor-burden estimation results. In Section 2.4, we discuss the convergence behavior of the proposed segmentation algorithm.

2.2 Materials and Methods

2.2.1 MRI

All studies were performed in accordance with the guidelines of the Washington University Animal Studies Committee and in accordance with protocols approved by the Washington University Division of Comparative Medicine that met or exceeded American Association for the Accreditation of Laboratory Animal Care standards. Respiratory-gated, spin-echo MR images of mice were collected with a small-animal MR scanner based on an Oxford Instruments (Oxford, UK) 4.7 tesla, 40-cm bore magnet. The magnet is equipped with Agilent/Magnex Scientific (Yarnton, UK) actively shielded, high-performance (21 cm inner diameter, ∼30 G/cm, ∼200 μs rise-time) gradient coils and International Electric Company (Helsinki, Finland) gradient power amplifiers and is interfaced with an Agilent/Varian NMR Systems (Santa Clara, CA) DirectDrive^{TM} console. All data were collected using a Stark Contrast (Erlangen, Germany) 2.5 cm birdcage-style rf coil. Prior to the imaging experiments, mice were anesthetized with isoflurane and were maintained on isoflurane/O_2 (1-1.25 % v/v) throughout data collection. Animal core-body temperature was maintained at 37 ± 1 °C by circulation of warm air through the bore of the magnet. During the imaging experiments, the respiration rates for all mice were regular and ∼2 sec^{-1}. Synchronization of MR data collection with animal respiration was achieved with a home-built respiratory-gating unit [21] and all images were collected during post-expiratory periods. Twenty-four contiguous coronal slices, ventral to dorsal, were collected for each mouse. Imaging parameters were TR ∼3 s, TE = 20 ms, FOV 2.5 cm × 2.5 cm^2, slice thickness = 0.5 mm, 128 × 128 data matrix, 4 averages. These scan parameters were chosen to maximize the contrast between healthy lung tissue and tumor.

2.2.2 Algorithm Development

Our algorithm for lung segmentation is based upon a 2D parametric lung-shape model. A 2D model was chosen over 3D modeling because (1) our data are composed of 2D MRI slices, and (2) the number of parameters to jointly optimize is fewer than in a 3D model, without

loss of fidelity. Simpler models, i.e., models with fewer parameters, are both more robust and are more efficiently optimized.

To fit our proposed parametric model to each torso slice, we propose an objective function, described in detail below, with which we can find locally optimal parameter values using the Nelder-Mead simplex method [46]. The Nelder-Mead simplex method is appropriate here because is it an unconstrained, non-linear optimization method for objective functions in high-dimensional spaces.

This section is organized as follows. First, we introduce our proposed parametric lung model. Next, we describe how our algorithm is initialized. Finally, we introduce our objective function and its components.

Parametric Model We introduce a lung model composed of four curves – parabolic segments AC and CB, and mixed parabolic segments AD and DB, shown in Figure 2.2.a. These curves are defined by their endpoints and by 6 curvature parameters – a_{AC}, a_{CB}, $a_{AD,1}$, $a_{AD,2}$, $a_{DB,1}$, and $a_{DB,2}$. The equation for any parabolic segment JK with curvature parameter a_{JK} and endpoints (x_J, y_J) and (x_K, y_K), with $x_J < x_K$, is

$$JK = \{(x,y) | y = a_{JK}x^2 + b_{JK}x + c_{JK}, \quad x_J \leq x \leq x_K\}, \tag{2.1}$$

where parameters b_{JK} and c_{JK} are defined by

$$b_{JK} = \frac{y_K - y_J - a_{JK}(x_K^2 - x_J^2)}{x_K - x_J} \tag{2.2}$$

$$c_{JK} = y_J - b_{JK}x_J - a_{JK}x_J^2. \tag{2.3}$$

The mixed parabolic segments are weighted sums of two such parabolic segments. For example, AD is defined as

$$\begin{aligned} AD = \{(x,y) | y = & \frac{x - x_A}{x_D - x_A}(a_{AD,1}x^2 + b_{AD,1}x + c_{AD,1}) + \\ & \left(1 - \frac{x - x_A}{x_D - x_A}\right)(a_{AD,2}x^2 + b_{AD,2}x + c_{AD,2}), \quad x_A \leq x \leq x_D\}. \end{aligned} \tag{2.4}$$

Because the center line of the lungs is also the approximate symmetry line of the overall lung shape, we reduce the number of parameters by setting $x_D = x_C$.

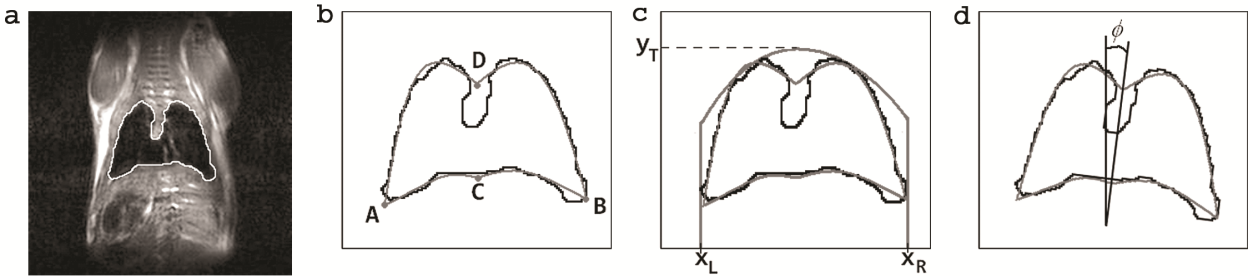

Figure 2.2: Illustration of fitted parametric model: (a) threshold-based segmentation of mouse lung slice from Figure 2.1.a; (b) parabolic segments AC, CB, AD, and DB; (c) mask defined by two vertical bounds at x_L and x_R and a parabola centered between the two bounds at y_T; and (d) rotation parameter ϕ.

To allow for a tighter fit over all images, we also impose a mask (Fig. 2.2.b), with edges defined by two vertical bounds, at x_L and x_R, and a parabola with curvature parameter a_{LR}, centered at (x_C, y_T). Only pixels that lie within both the lung model and this mask are classified as lung. In addition, we include a rotation parameter ϕ, shown in Figure 2.2.c, which allows for variation in mouse position within the imaging FOV. Due to contraints on how the mouse can be placed within the scanner, for our data the rotation parameter varies roughly between -10° and +10°. Thus, the full set of parameters defining the lung shape, θ, is defined as

$$\theta = \{x_A, y_A, x_B, y_B, x_C, y_C, y_D, a_{AC}, a_{CB}, a_{AD,1}, a_{AD,2}, a_{DB,1}, a_{DB,2}, x_L, x_R, a_{LR}, y_T, \phi\} \qquad (2.5)$$

Initialization We manually initialize our algorithm with a rough segmentation of one interior MRI slice. We fit our model parameters to the manual segmentation using the simplex optimization method, maximizing the overlap of the manual segmentation, S_{man}, and the parametric segmentation, $S(\theta)$.

$$\text{Overlap}\left(S_{\text{man}}, S(\theta)\right) = \frac{|S_{\text{man}} \cap S(\theta)|}{|S_{\text{man}} \cup S(\theta)|} \qquad (2.6)$$

The parameters are propagated forward and backward from the chosen starting slice and used as initializations for the adjacent slices. Using the simplex method, we fit the model parameters to the new slices by minimizing an objective function, which is discussed below. The optimized parameters are then propagated to the next adjacent slices, until the segmentation is complete.

Objective Function To fit our model parameters to the current slice, we minimize an objective function, $O(\theta)$, that is a sum of several 'goodness of fit' metrics: an intensity-weighted overlap metric, O_{weight}; the Manhattan distance between the interior and exterior voxel distributions, $O_{\text{distribution}}$; a measure of the magnitude and direction of changes in parameter values from the adjacent slice to the current slice, O_{change}; and the symmetry and concavity of the model given the current parameter values, O_{shape}. We define $O(\theta)$ as

$$O(\theta) = O_{\text{weight}}(\theta) + O_{\text{distribution}}(\theta) + \sum_n O_{\text{change}}(\theta_n) + \sum_m O_{\text{shape}}(\theta_m), \tag{2.7}$$

where O_{weight} and $O_{\text{distribution}}$ take values between zero and one, and O_{change} and O_{shape} take values greater than or equal to zero.

O_{weight} We want to reward inclusion and exclusion of voxels based on the likelihood that the intensities came from lung or non-lung, given the manual initialization. To this end, we first construct a new image with intensity $\mathbf{W}$.

Let $\mathbf{I}(x, y)$ be the image intensity at voxel (x, y) of the current slice, as in Figure 2.3.a. We fit a cubic polynomial p, shown in Figure 2.3.b, to the difference in histograms of the manually-segmented lung and non-lung.

$$\mathbf{W}(x, y) = \begin{cases} p(\mathbf{I}(x, y)) & \text{if } |p(\mathbf{I}(x, y))| > 0.2 \\ 0 & \text{else} \end{cases} \tag{2.8}$$

In this new image $\mathbf{W}$, as in Figure 2.3.c, most positive areas should be within the lung segmentation, while most negative areas should be outside the segmentation.

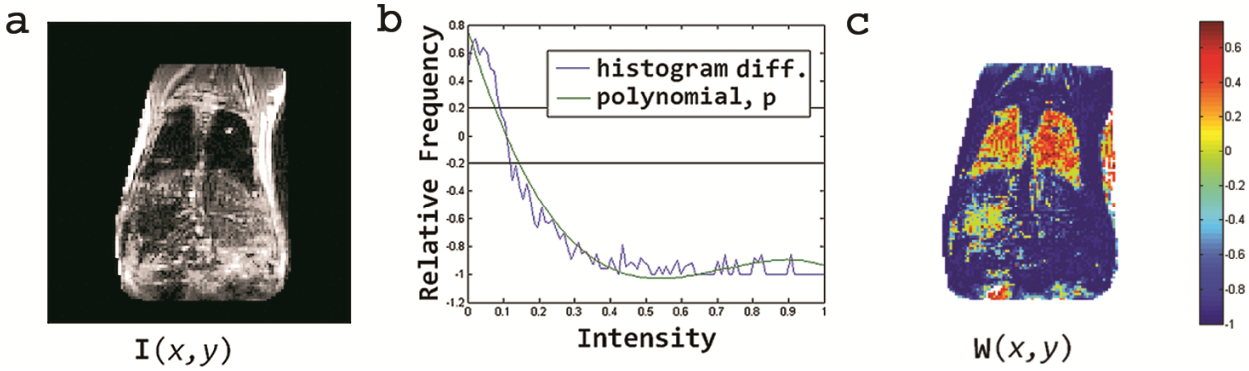

Figure 2.3: Illustration of the construction of weight image $\mathbf{W}$: (a) example original image $\mathbf{I}$; (b) graph of the difference between the histogram of manually-segmented lung and that of non-lung, along with the fitted cubic polynomial p; and (c) the new weight image $\mathbf{W}$.

Let θ be the current set of parameter values, $\mathbf{M}_{\text{in}}(\theta)$ the set of voxels within the current parametric lung segmentation and $\mathbf{M}_{\text{out}}(\theta)$ the set outside the segmentation. Note that $\mathbf{M}_{\text{out}}(\theta)$ does not include 'background,' voxels outside the body of the mouse, which are removed as a preprocessing step. We define $O_{\text{weight}}(\theta)$ as

$$\begin{aligned} O_{\text{weight}}(\theta) =& 0.75 \left(1 - \frac{\sum_{(x,y)\in\mathbf{M}_{\text{in}}(\theta)} \mathbf{W}_+(x,y)}{\sum_{\text{all } (x,y)} \mathbf{W}_+(x,y)} \right) \\ &+ 0.25 \left(1 - \frac{\sum_{(x,y)\in\mathbf{M}_{\text{out}}(\theta)} \mathbf{W}_-(x,y)}{\sum_{\text{all } (x,y)} \mathbf{W}_-(x,y)} \right), \end{aligned} \tag{2.9}$$

where we define $\mathbf{W}_+$ and $\mathbf{W}_-$ as

$$\mathbf{W}_+(x,y) = \begin{cases} \mathbf{W}(x,y) & \forall\ (x,y) : \mathbf{W}(x,y) > 0 \\ 0 & \text{else} \end{cases} \tag{2.10}$$

$$\mathbf{W}_-(x,y) = \begin{cases} \mathbf{W}(x,y) & \forall\ (x,y) : \mathbf{W}(x,y) < 0 \\ 0 & \text{else} \end{cases}. \tag{2.11}$$

Thus, $O_{\text{weight}}(\theta)$ is zero when all positive voxels in $\mathbf{W}$ lie within the parametric segmentation and all negative voxels lie outside the segmentation. Note that the weightings for the two components of $O_{\text{weight}}(\theta)$ in (2.9) are unequal. The first component rewards inclusion within the lung segmentation of voxels with 'lung-like' intensities. The second rewards exclusion from the segmentation of voxels with 'non-lung' intensities. The first component is weighted more heavily because, due to noise and partial-volume effects, it is expected that some voxels within the lungs will have 'non-lung' intensities, while extra-lung tissues are, in general, less likely to resemble lung tissue. However, the final segmentation is relatively insensitive to the choice of these weightings - comparison of segmentations found using the 0.75/0.25 weightings and equal 0.5/0.5 weightings had an average overlap of 94%.

$O_{\text{distribution}}$ Because our approach is based on differing intensity distributions inside and outside the lung, we also directly compute this difference in distributions, rewarding large differences. We define $O_{\text{distribution}}(\theta)$ as

$$O_{\text{distribution}}(\theta) = 1 - 0.5 \sum_{n=1}^{N} |\mathbf{h}_{\text{out}}(\theta) - \mathbf{h}_{\text{in}}(\theta)|, \tag{2.12}$$

where $\mathbf{h}_{\text{in}}(\theta)$ is a normalized histogram of the voxel intensities in $\mathbf{M}_{\text{in}}(\theta)$, and $\mathbf{h}_{\text{out}}(\theta)$ is a normalized histogram of the voxel intensities in $\mathbf{M}_{\text{out}}(\theta)$. Hence, $O_{\text{distribution}}(\theta)$ is equal to one when the histograms inside and outside the parametric segmentation match exactly, and its value decreases to a minimum of zero as the difference between the histograms increases.

O_{change} Since adjacent lung slices must form a continuous 3D lung surface, we penalize parameters with large changes from one slice to the next. For each parameter θ_n, with value $\theta_{n,0}$ in the adjacent slice, we define $O_{\text{change}}(\theta_n)$ as

$$O_{\text{change}}(\theta_n) = \begin{cases} 0 & \text{if } |\theta_n - \theta_{n,0}| < c \\ |\theta_n - \theta_{n,0}| & \text{else} \end{cases} \tag{2.13}$$

where c is a threshold on the magnitude of the parameter change from one slice to the next.

To enforce proper relative size of slices, we include an additional term in $O_{\text{change}}(y_C)$ and $O_{\text{change}}(y_D)$ that penalizes expansion of the lungs as the algorithm progresses toward the back and contraction of the lungs as it progresses toward the chest. For example, we define $O_{\text{change}}(y_C)$ as

$$O_{\text{change}}(y_C) = \begin{cases} 0 & \text{if } |y_C - y_{C,0}| < c \\ |y_C - y_{C,0}| & \text{else} \end{cases} + \begin{cases} 0 & \text{if } y_C < y_{C,0} \\ y_C - y_{C,0} & \text{else.} \end{cases} \tag{2.14}$$

O_{shape} In general, the boundary of the lungs is roughly left/right symmetric, so we penalize large asymmetries in the fitted lung shape. For each of the three pairs of curvature parameters a_i and a_j, where a_j is the corresponding value for a_i from the opposite side of the lung, we define $O_{\text{shape,sym}}(a_i, a_j)$ as

$$O_{\text{shape,sym}}(a_i, a_j) = \begin{cases} 0 & \text{if } |a_i - a_j| < c \\ |a_i - a_j| & \text{else} \end{cases} \tag{2.15}$$

where c is a threshold on the asymmetry of the segmentation. Similarly, each curvature parameter contributes a term $O_{\text{shape,conc}}$ that penalizes convexity of the lung curves.

2.2.3 Pathology

Two tumor cell lines, B16 murine malignant melanoma from the ATCC (Manassas, VA) and WT9614 3-methylchlantherene fibrosarcoma (kindly provided by Robert Schreiber, Washington University in St. Louis), were injected intravenously into age matched C57Bl6 male mice at 2.5x10^5 cells per animal. These animals were sacrificed, along with age- and sex-matched, saline-injected control mice, at various points after tumor injection ranging from 10 days to three weeks. Upon sacrifice, the lung block was removed through a sternotomy and trimmed free of the mediastinal tissue, leaving only lung parenchyma and airways attached. The tumor bearing lung block was weighed (Series 320 XT Analytical balance, Precisa, Golden CO) and the total tumor burden calculated by subtracting the weight of non-tumor bearing control lungs from that of the tumor bearing lungs.

2.2.4 Evaluation Criteria

To validate the performance of our segmentation method, we first show that our automatic lung segmentation is comparable to that of expert human segmenters. We then show that the average intensities of the segmented lungs in both the automatic and manual segmenter cases correlate well with the tumor burden measured using lung weight. All validation studies were conducted using only data sets that were not used for algorithm development.

Manual segmentations of the lungs were generated independently by four experts for 6 of the 27 imaged mice. These mice were selected in an unbiased manner so as to cover, as uniformly as possible, the full spectrum of tumor burdens present in the data.

To compare segmentations from two different segmenters, we use the following overlap metric:

$$\text{Overlap}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cap \mathbf{B}}{\mathbf{A} \cup \mathbf{B}}, \tag{2.16}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the two sets of voxels designated as lung by the two segmenters. This metric is useful in this case because we do not have a ground truth segmentation and, therefore, cannot use a metric like percent error.

2.3 Results

In this chapter, we present a method for accurate and reproducible lung segmentation in mice with heavy and/or diffuse tumor (Fig. 2.4). This method allows nearly fully automatic measurement of tumor burden in the lungs. Table 2.1 shows the average overlap between each pair of independently drawn manual segmentations. Table 2.2 shows the total average overlap of our automatic segmentation results with each of the manual segmenters. In all the results, the initializations for the automatic segmentation were generated by an additional segmenter, independently from the expert segmentations. As can be seen from the tables, there is generally good agreement amongst the results for human segmenters, as well as between the human segmenters and the automatic result, in terms of which areas in the MR images are classified as lung.

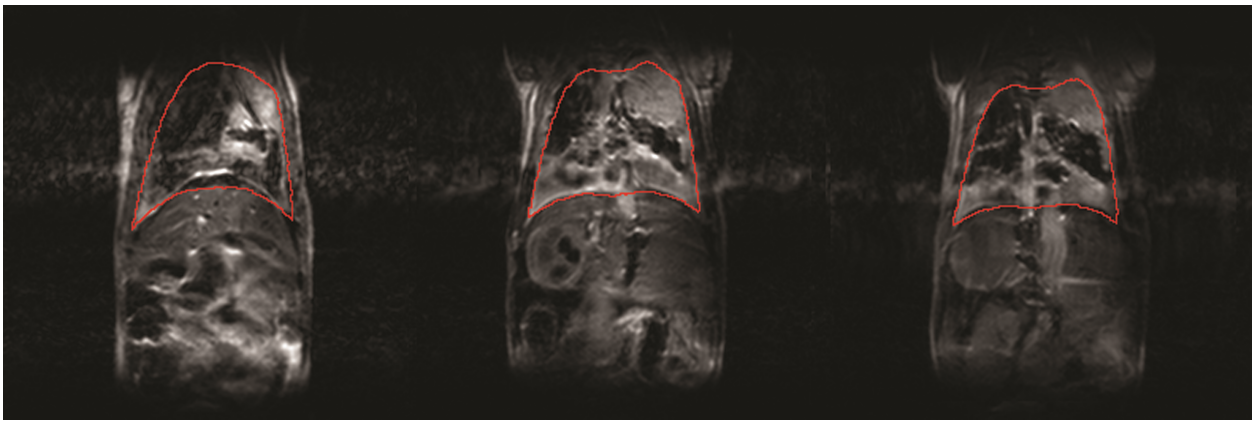

Figure 2.4: Three example MRI slices with results from the automatic MRI tissue segmentation method outlined in red.

Table 2.1: Average Percent Overlap between the Four Manual MRI Tissue Segmentations

Expert	A	B	C	D
A	100	84.70 ± 6.34	74.39 ± 9.66	79.33 ± 7.69
B		100	76.30 ± 8.03	82.52 ± 7.56
C			100	75.61 ± 8.44
D				100

Values are given as the mean $\pm$ standard deviation

Table 2.2: Average Percent Overlap between the Automatic MRI Tissue Segmentation and the Four Manual Segmentations

Expert	A	B	C	D
	76.03 ± 3.72	78.72 ± 2.61	68.87 ± 8.68	71.97 ± 7.69

Values are given as the mean $\pm$ standard deviation

Because the goal is to quantify tumor via image intensity, a fairer metric of the correspondence between two segmentations may be derived by comparing the average image intensities within the two segmentations. Table 2.3 shows the average percent difference in intensity between each pair of independently drawn manual segmentations. Table 2.4 shows the total

percent difference in intensity between our automatic segmentation results and each of the manual segmenters. As these tables show, there is good agreement in the average image intensities derived from the expert manual segmenters and the automatic result.

Table 2.3: Average Percent Difference in Intensity between the Four Manual MRI Tissue Segmentations

Expert	A	B	C	D
A	0	5.06 ± 6.46	19.24 ± 15.49	8.34 ± 5.76
B		0	14.86 ± 10.67	9.38 ± 6.78
C			0	16.34 ± 10.54
D				0

Values are given as the mean ± standard deviation

Table 2.4: Average Percent Difference in Intensity between the Automatic MRI Tissue Segmentation and the Four Manual Segmentations

Expert	A	B	C	D
	12.96 ± 11.29	10.69 ± 9.32	20.77 ± 10.15	16.29 ± 14.53

Values are given as the mean ± standard deviation

The key validation of our method is the correlation between total tumor burden measured by lung weight and the average intensities of the manual and automatic results. Table 2.5 shows the correlation of each segmenter's computed average intensities with the lung weights of the six mice. Outlier intensities were present both in the full set of mice and in the subset of 6 manually segmented mice. We noted that in these outlier images, the overall intensity of the images for a particular animal was either darker or brighter than the average image in the set. To correct for this variation, lung-image intensities were normalized based upon the image intensity of the liver with the same mouse. The liver intensity was calculated by manual selection of a region of interest (ROI) containing only liver. Outlier voxels, which may be due to liver tumor or liver vasculature, with intensities beyond one standard deviation from the mean intensity, were automatically discarded prior to computing the average intensity within the ROI. Table 2.6 shows corrected correlations, in which the same normalizing liver

intensities were used for each segmenter's average intensities. The results shown in these tables demonstrate an excellent correlation between these corrected lung intensities and tumor burden, as measured by lung weight. The correlation between the corrected average intensities found by the automatic method and the lung weights for the complete set of 27 mice was 0.93 (Fig. 2.5.a). Bland-Altman analysis of the lung weight and the automated corrected average intensities shows that the limits of agreement are 0.3mg $\pm$ 168.9 (defined as the bias $\pm$ 1.96 times the standard deviation of the difference). A plot of this analysis is shown in Figure 2.5.b.

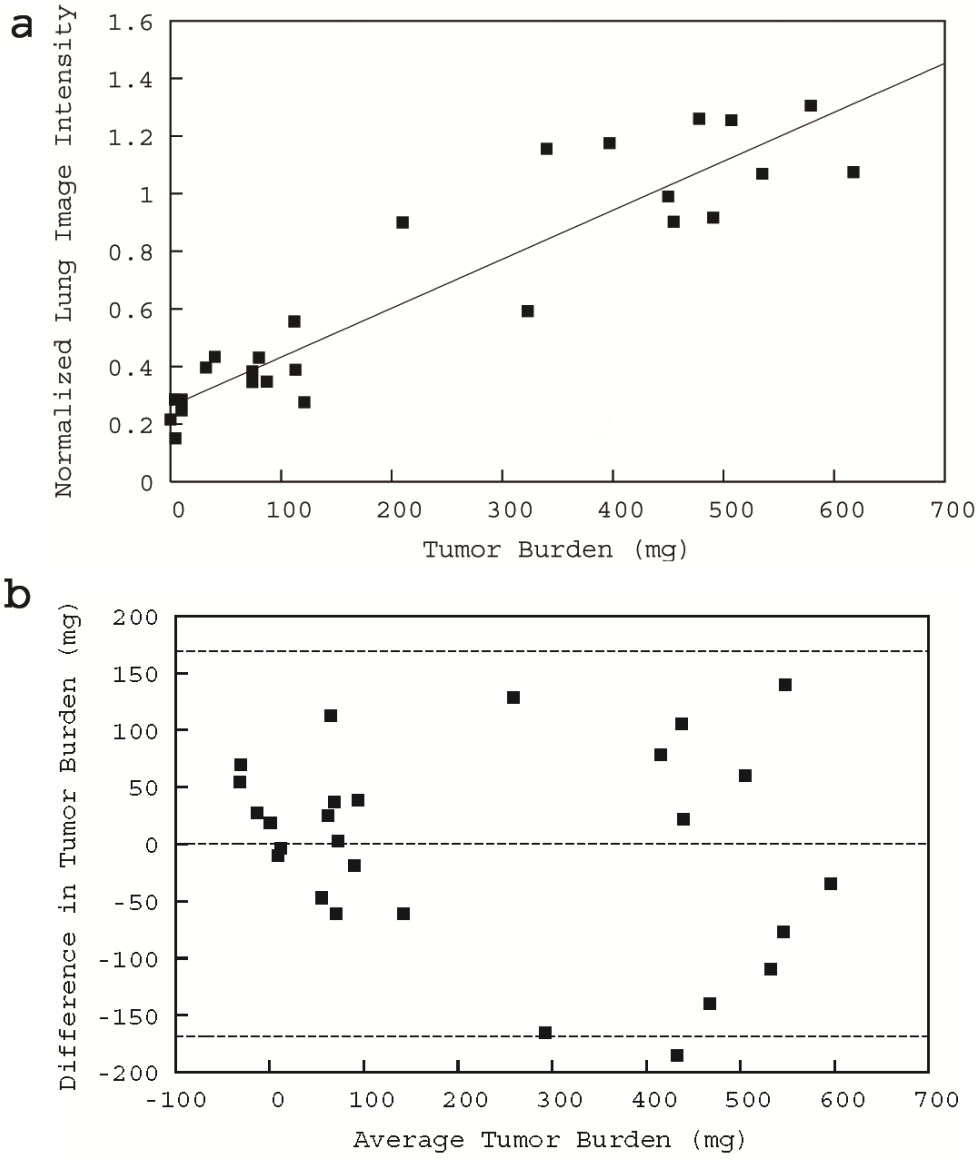

Figure 2.5: (a) Plot of corrected average intensity versus tumor burden for the automatic MRI tissue segmentation method, with the least squares linear fit ($R^2 = 0.86$) shown. For each mouse, lung-image intensities were normalized based upon the corresponding liver-image intensity for that mouse. (b) Bland-Altman plot for the same data, where average intensity has been converted into mg tumor following the least squares linear regression in (a).

Table 2.5: Correlation between Average Intensities and Lung Weights

Automatic	Expert A	Expert B	Expert C	Expert D
0.765	0.776	0.766	0.749	0.741

Table 2.6: Correlation between Corrected Average Intensities and Lung Weights

Automatic	Expert A	Expert B	Expert C	Expert D
0.955	0.965	0.962	0.932	0.943

In summary, Table 2.2 demonstrates the excellent congruence between our automated lung segmentations and those generated manually by a panel of four experts in a series of six mice. As reported in this table, the percent overlap of lung pixels amongst automated and manual segmentations ranges from 72.0 to 78.7%, compared to a range of 79.3 to 84.7% between manual segmentations. The identification of tumor in lung is dependent upon the relatively bright image intensity of tumor compared with healthy lung tissue. As shown in Table 2.4, differences in average lung-image intensity between automated and manual segmentations in these same mice are relatively small, ranging from 10.7 to 16.3% across the panel of segmenters. Finally, the correlation between average lung image intensity generated by automated lung segmentation and measured lung weight is greater than 0.93 (Fig. 2.5.a), which corresponds to a coefficient of determination greater than 86%, demonstrating clearly that average lung image intensity provides a useful measure of tumor burden in lung with diffuse or heavy tumor burden.

2.4 Discussion

2.4.1 Segmentation Convergence

Despite the large number of parameters to be optimized, 18 per slice, and the non-convexity of the optimization formulation, the proposed algorithm converges to an accurate full lung

segmentation very reliably, with little sensitivity to the manual initialization. This low initialization sensitivity is primarily due to the fact that small changes in initial segmentation result in very small changes in the intra-lung and extra-lung voxel intensity distributions.

Conversely, the segmentation algorithm is somewhat sensitive the the symmetry axis of the manual segmentation. The versatility of the mouse-lung model proposed herein allows for a tight fit of the parametric model to a manual segmentation even with a relatively large error in the rotation parameter ϕ. However, if ϕ differs sufficiently from the angle of the true symmetry axis, as the segmentation progresses beyond the initial slice, the segmentation will continuously degrade. This degradation occurs because the asymmetry penalty is predisposing the optimized parameters to symmetry along an incorrect axis. Thus, given a sufficiently poor initialization, moving from the initial slice toward the mouse's chest and back, the segmentations will generally become less and less accurate.

In theory, any slice can be used for the manual initialization. In practice, the best results are obtained when initializing with a slice from the center of the mouse. There are two reasons for this. First, using a central mouse slice minimizes the physical distance that the automatic segmentation algorithm moves away from the manual segmentation. Second, slices at the extremes of the lung (far into the chest or the back) exhibit shape characteristics that make them poor choices for initialization. For instance, near the chest of the mouse, the two curves forming the upper lung boundary blend into a single curve, as seen in Figure 2.4. Accurate segmentation of these MR slices can be achieved by adjustment to the curvature parameters of AD and DB or via the additional mask, by adjusting a_{LR} and y_T. Therefore, only one of these two parameter sets must have a 'good' value for the parametric segmentation to match the initializing manual segmentation well. Given the likelihood of 'bad' initial values for the other parameter set, manual initialization using such a slice is unlikely to yield accurate segmentations.

2.4.2 Impact

In Section 2.3, we showed that the results of the proposed automatic segmentation algorithm are on par with expert manual segmentation. Given the reliability of the automatic results, two benefits are apparent. First, the automatic method can be used in place of manual

segmentation in order to reduce the time and effort currently spent by experts on manual segmentation. Second, the results from the automatic segmentation method can be used to generate a calibration curve with which the MR image intensity metric for tumor burden can be converted into an absolute measure of tumor burden, e.g., mg of tumor. This second point is important because, as shown in Tables 2.6 and 2.3, the expert segmentation-derived tumor burdens each correlated well with the tumor weight, but they were inconsistent with each other. Because of this, it would not be advisable to use one expert's results to generate a calibration curve, and then use that curve to convert a second expert's results from average intensity to mg tumor. Having an automatic method removes the difficulty of requiring one expert to segment all the MR images.

2.5 Conclusion

The use of mouse models to aid in the development and monitoring of new therapies for lung cancer requires the ability to accurately measure lung tumor burden *in vivo*. In this chapter, we have demonstrated that corrected average MR image intensity in mouse lung is an accurate metric of total tumor burden. The tumor measurements were validated by correlating MR image intensities with the weight of the excised lungs. By measuring average MR lung intensity, tumor burden can be estimated *in vivo*, even in cases of diffuse disease where individual tumors cannot be segmented from the MR images. Thus, *relative* measures of tumor burden for a single animal can be established simply by comparing average lung intensities from images collected at different time points. As described herein, *absolute* tumor burden measures can also be determined following establishment of a calibration curve between MR image intensities and lung weights. Because this average image intensity approach requires accurate lung segmentation, efficiency and throughput of analysis would be greatly improved through use of an automated segmentation routine.

We have described a novel method for automated segmentation and analysis of the MR images of murine lungs and pulmonary tumors. We have developed a new, two-dimensional parametric model for mouse lung that accurately preserves the overall shape of the lungs, and a novel cost function for optimization of the model parameter values for each lung image. Qualitatively, our segmentation results are well fitted to the lungs. Quantitatively,

the correlation between the corrected average intensity tumor-burden metric and the lung weight is excellent, and comparable to that of fully-manual expert segmentation.

Future work includes fully automating the segmentation software to further improve throughput. The lung segmentation method could be successfully adapted to other problems where parametric models are of use, including prostate cancer analysis.

2.6 Acknowledgements

This work was supported by a Mr. and Mrs. Spencer T. Olin Fellowship for Women in Graduate Study; NSF Grant CCF-0963742; the NIH/NCI Small Animal Imaging Resource Program (U24 CA83060); the Alvin J. Siteman Cancer Center at Washington University in St. Louis, an NCI Comprehensive Cancer Center (P30 CA91842); an NIH/NCI Grant (KO8 CA131097); and an American Thoracic Society/Lungevity Foundation Research Grant.

Chapter 3

Automatic Segmentation of Rodent Spinal Cord Diffusion MR Images

---

This chapter was previously published in *Magnetic Resonance in Medicine*:
V. K. Tidwell, J. H. Kim, S.–K. Song, and A. Nehorai. Automatic segmentation of rodent spinal cord diffusion MR images. *Magnetic Resonance in Medicine*, 62:893–901, 2010.

I designed, implemented, and tested the automatic segmentation algorithm. J.H.K. and I performed the data analysis and wrote the manuscript. J.H.K. and S.K.S. designed the imaging experiments and collected the MR images; A.N. helped with data analysis.

MRI is a key tool for noninvasive spinal cord lesion analysis; however, accurate, quantitative methods for this analysis are lacking. A new, multistep, multidimensional approach, utilizing the Classification Expectation Maximization algorithm, is proposed for MRI segmentation of spinal cord tissues. Diffusion tensor imaging is used to generate multiple images of each spinal slice, with different diffusion direction weightings. The maximum likelihood tissue classifications are then jointly estimated to produce a binary classification image, corresponding to voxels containing either spinal cord or background. Edge detection is employed to find a nonparametric curve encapsulating the entire spinal cord. The algorithm is evaluated using data from *in vivo* diffusion tensor imaging of control and injured mouse spinal cords. The algorithm is shown to remain accurate for whole spinal cord, white matter, and

hemorrhage segmentation in the presence of significant injury. The results of the method are shown to be at least on par with expert manual segmentation.

3.1 Introduction

In spinal cord injury, the amount of total parenchyma or surviving white matter is known to be strongly related to post-injury neurological function [51, 40, 2, 1]. Objective quantification of these regions of interest is critical in both fundamental pathophysiological study and the development of effective treatment. The most universally accepted method for accurate segmentation analysis is histology [43, 41, 3], but its use is limited to postmortem study due to its invasive nature. In contrast, MRI is well suited for noninvasive diagnosis of living tissue. MRI-based spinal cord lesion reporting, both *in vivo* and *ex vivo*, shows good agreement with conventional histology validation and reflects clinical disabilities [32, 39, 10, 52]. However, there is a lack of objective and precise quantitative methods for segmentation of total parenchyma or white matter in MR images.

Existing methods for *in vivo* transaxial spinal cord segmentation in MRI can generally be categorized into two broad classes. The first, most common class requires significant human intervention, and ranges from entirely manual segmentation to computer-aided manual edge selection. These segmentation methods are subject to human bias and are therefore unreliable and generally not reproducible. They are also slow and therefore impractical for analyzing large data sets. The second class seeks to define a contour around the cord automatically, based on image gradients and pixel intensities within and outside the contour, with minimal human intervention. These approaches generally use contour methods such as snakes [30] or level-sets [47], and they vary in speed but are in general much faster than manual segmentation. A recent example of this second class is presented in Deng et al. [14], which uses a B-spline snake approach to find the spinal cord contour from *in vivo* MR images of healthy and mildly injured rat spines. This method relies on human intervention to select the midpoint of the spine as the seed point for the snake algorithm, and it generates segmentations in an average of 1.6 seconds per slice [14].

The automatic methods outlined above all make assumptions about the cord shape that do not necessarily hold in the case of an injured spinal cord. Injured cords shrink as the tissue

atrophies and can assume very irregular shapes. Contour-based methods such as snakes or manual edge tracing assume relatively smooth edges. Additionally, severely injured cords may have voids or hemorrhage within the spinal cord, which violates the basic assumption of contour-based methods—that a single continuous boundary can be found to separate the tissue of interest from the rest of the image. Robust methods that eliminate these faulty assumptions are needed.

We propose a multidimensional, multistep Classification Expectation Maximization (CEM)–based algorithm for spinal cord segmentation. Our algorithm is multidimensional in that it segments the cord based on a set of MR images collected for diffusion tensor imaging (DTI) analysis. This joint segmentation incorporates significantly more data than is available in a single MR image and thus is more robust to noise in individual images.

We also extend our algorithm to automatic *in vivo* segmentation of spared white matter and regions of hemorrhage in injured spines. Previous studies of automatic MRI segmentation of spinal white/gray matter have only attempted to validate their algorithms for uninjured cords [16, 17], or excised cords, imaged *ex vivo* [61]. To our knowledge, no previous work has been done on automatically segmenting areas of spinal hemorrhage.

Our method is novel in medical image segmentation due to its multistep approach, which allows improved segmentation accuracy by incorporating both prior knowledge of cord geometry and the distinct information contained in the different images in successive steps. For example, in the proposed algorithm, the initial step for rough spinal cord segmentation is based on the b=0 image, which is acquired without diffusion-sensitizing gradient pairs and so in general is a T_2-weighted (T2W) image. The next steps provide further, more exact, segmentation based jointly on the diffusion-weighted images (DWIs). To our knowledge, no previous work on automatic *in vivo* transaxial MRI spinal cord segmentation has incorporated the information in T2W images in addition to the DWIs [14, 16, 61].

Additionally, to our knowledge the CEM algorithm has not been used for spine segmentation. Our algorithm differs greatly from previous, contour-based cord segmentation approaches in that our algorithm is a voxel classification algorithm—voxels are classified individually rather than grouped according to a single contour. In contrast to existing contour-based algorithms, the proposed algorithm defines contours only as a means to generate localization constraints on which voxels may be classified as particular tissue types.

In Section 3.2, I present the challenges for injured spinal cord segmentation. In Section 3.3, I introduce the proposed CEM-based segmentation algorithm. In Sections 3.4 and 3.5, I present and then discuss results for both manual and automatic segmentation of MR images of injured spinal cords.

3.2 Background

Spinal cord histology allows high resolution tissue segmentation with clear separation between the tissue types. The goal for MRI spinal cord segmentation is to use *in vivo* imaging to approximate, as closely as possible, the tissue types that one could generate via histology, thus aiding the treatment and evaluation of spinal cord injuries. MRI is, however, much lower in resolution, and therefore generating these accurate segmentations is nontrivial. In this section, we discuss the relevant attributes of spinal cord diffusion MR images, and the challenges faced when automatically segmenting these images, particularly in the case of injured cords.

3.2.1 Characteristics of DTI

DTI provides microstructural information with greater sensitivity to tissue integrity than conventional MRI. Many studies have reported the potential of six-direction DTI-derived parameters to reveal the morphological integrity and pathophysiological changes of living tissue in rodent spinal cord studies [32, 5, 11, 26].

In the T2W image, the brightest area is located within the spinal canal, containing both cord and cerebrospinal fluid (CSF), with minimal differentiation in voxel intensity level between the white and gray matter tissue types. In severely injured cords, the intensity level of the CSF can on occasion differ slightly from that of the cord; however, even when there is variation between the cord and CSF, the difference in intensity level between the background and the spinal cavity is always much greater.

DWIs are brightest only in the region corresponding to spinal cord, not CSF. The voxel intensity level of the CSF in the DWIs is equivalent to that of the background tissues. This

is because the signal from CSF attenuates significantly in diffusion weighted imaging in any direction. Voxel intensities in DWIs differentiate strongly between white and gray matter; in some diffusion gradient directions, the white matter is brighter than the gray matter, while in other directions, the gray matter is brightest.

From the DWIs, DTI maps can be calculated. These DTI maps—in particular, the relative anisotropy, the axial diffusivity, and the radial diffusivity—do not have the same useful properties as the T2W image or the DWIs to allow for simple segmentation of the spine from the background tissue. However, they have clearer intensity separation between the white and gray matter than the DWIs, particularly in the case of injured cords.

3.2.2 Challenges for Automatic Segmentation

There are many challenges for accurate voxel-by-voxel classification of the spinal MR images. For instance, there are frequently scattered bright spots in both the T2W image and in the DWIs that lie outside the spinal cord. Because the two types of images are sensitive to different tissue properties (spin-spin relaxation and diffusion), these extraneous bright spots are in general not colocated in the T2W image and the DWIs. Thus, by using both types of images, we can achieve more accurate voxel-by-voxel classification of the tissues.

Additionally, there are occasionally colocated bright areas in both the T2W image and the DWIs, e.g., at points where nerves branch off from the spinal cord. Because such nerve tissue is similar to the cord tissue, it cannot be classified as background by voxel intensity alone. Therefore, accurate segmentation requires localization constraints in addition to the pure voxel classification of the CEM algorithm.

Finally, in assessing chronic posttraumatic changes in injured cords, we observe two major effects of injury—atrophy and hemorrhage. As the cord atrophies, the size of the spinal cord decreases, making the shape of the injured cord, as well as white and gray matter areas, unpredictable. The additional space within the spinal cavity is filled by CSF. In addition, the tissue contrast of the surviving white and gray matter in injured cords is much less clear than in the control cords.

In injured cords, the presence of hemorrhage appears as dark areas within the bright spinal cord. Hemorrhage has a similar voxel intensity distribution to background, non-cord voxels in the MR images because its relaxation and diffusion properties differ from those of the spinal cord, for example, the $T_2\star$ value is shorter in hemorrhage. It is useful to automatically produce two spinal cord segmentations, one including hemorrhage and one excluding it. This allows quantitative analysis of the size of the hemorrhage relative to the whole cord size, which is useful in spinal cord injury evaluation.

By using a multistep approach to the automatic spine segmentation problem, we are able to take advantage of the unique tissue differentiation abilities of the MR image types. Our multidimensional approach allows us to generate segmentations that are more robust to noise in individual images than existing approaches that rely on a single MR image.

3.3 Methods

3.3.1 Theory: Classification Using the CEM Algorithm

We propose using intensity levels for automatic segmentation rather than the existing contour-fitting approaches, to allow more accurate segmentation of injured spinal cords. We assume the intensity values in the T2W image come from a sum of two distinct Gaussian distributions—that of the background and that of the cord and CSF. Similarly, we assume the intensity values in each DWI come from a sum of three distinct Gaussian distributions—those of the background (all non-spinal cord tissues, including bone), the white matter, and the gray matter (Fig. 3.1). Although we make the assumption that the above distributions are Gaussian for the purposes of our classification algorithm, these distributions are known to be Rician. We also compute results under a Rician assumption for purposes of comparison.

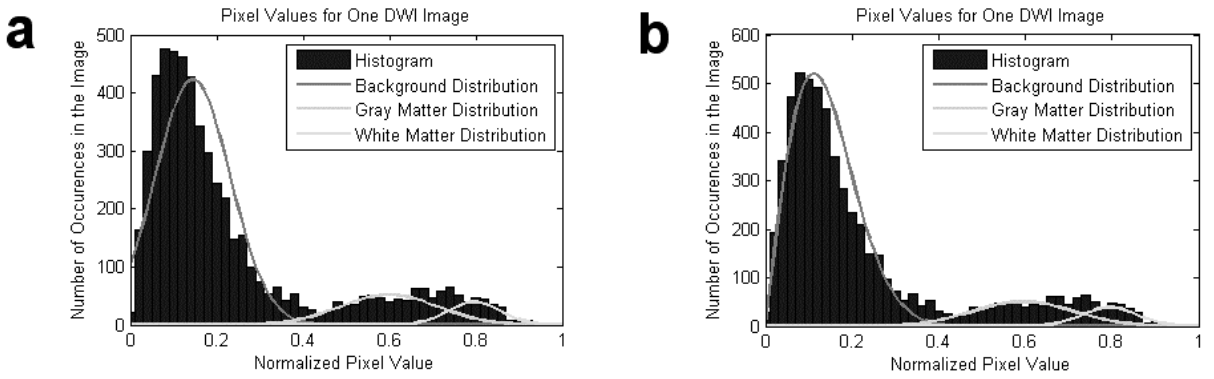

Figure 3.1: Histogram of a representative cropped and normalized DWI for a control spine (see image in Figure 3.2.b). The distributions of the automatically segmented tissue types (background, white matter, and gray matter) are superimposed assuming data comes from (a) Gaussian and (b) Rician distributions.

To find an optimal, unbiased separation of voxels into these classes, we employ the CEM algorithm in a series of stages. Each stage of our algorithm separates the voxels into one of two classes, e.g., initially, the pixels are classified as either background or spinal cord, and then in a later stage, the cord voxels are classified as either spared white or gray/injured white matter.

We observe the vector of image intensity values $\mathbf{x}_i \in \mathbb{R}^D$ from voxels $i = 1, \ldots, n$ in D input images. These intensity values come from K possible tissue classifications $k = 1, \ldots, K$, where K is known. Let y_i denote the classification corresponding to $\mathbf{x}_i$ $(i = 1, \ldots, n)$, taking a value from 1 to K. The intensity distribution, $\mathbf{f}_k(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, for each class k is assumed to be a multidimensional Gaussian with mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$ both unknown. Each classification has prior probability π_k, which is also assumed to be unknown. We assume noise independence within an individual voxel across the T2W image and the six DWIs, and across the three DTI maps used; this reduces $\boldsymbol{\Sigma}_k$ to a diagonal matrix. Therefore, the estimated distribution parameters are reduced to the priors π_k, means $\boldsymbol{\mu}_k$, and variances $\boldsymbol{\sigma}_k$ of multidimensional Gaussian intensity distributions.

The CEM algorithm [8] is a variation of the EM algorithm [13], with a classification step (C-step) added between the expectation step (E-step) and the maximization step (M-step).

It finds a classification maximum likelihood estimate of y_i, maximizing the classification likelihood, C_l:

$$C_\text{l} = \sum_{k'=1}^{K} \sum_{i|y_i=k} \log \mathbf{f}(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{3.1}$$

Starting from initial cluster parameter values, the CEM algorithm iteratively converges to a maximum *a posteriori* estimate of y_i, which is known to be in general initialization-sensitive. For example, if the initial cluster parameter values are far from the optimal parameters, the estimate for y_i may be locally, not globally, optimal. To avoid such 'bad' initialization, for each CEM initialization in our method, we apply the commonly used K-means algorithm with random sample seeding to find a preliminary clustering of the data. For our application, 30 random K-means seedings produced exactly identical final spine segmentations for each MRI data set.

Given our assumptions, the steps of the CEM algorithm for iteration m are as follows [8], assuming Gaussian distributions:

- E-step: Compute the current posterior probabilities $p_{i,k}$ for all data points $i = 1, \ldots, n$ and all clusters $k = 1, \ldots, K$

$$p_{i,k} = \frac{\pi_k^{(m)} \mathbf{f}(\mathbf{x}_i|\boldsymbol{\mu}_k^{(m)}, \boldsymbol{\Sigma}_k^{(m)})}{\sum_{k'=1}^{K} \pi_{k'} \mathbf{f}(\mathbf{x}_i|\boldsymbol{\mu}_{k'}^{(m)}, \boldsymbol{\Sigma}_{k'}^{(m)})}. \tag{3.2}$$

- C-step: Assign each data point $\mathbf{x}_i$ to the cluster with the largest posterior probability.

$$y_i^{(m+1)} = \arg\max_k \, p_{i,k}. \tag{3.3}$$

- M-step: Compute the maximum likelihood estimates for the parameters, $\pi_k^{(m+1)}$, $\boldsymbol{\mu}_k^{(m+1)}$, and $\boldsymbol{\Sigma}_k^{(m+1)}$:

$$\pi_k^{(m+1)} = \frac{\sum_{i=1}^{n} \hat{p}_{i,k}}{n}, \tag{3.4}$$

$$\boldsymbol{\mu}_k^{(m+1)} = \frac{\sum_{i=1}^{n} \hat{p}_{i,k} \mathbf{x}_i}{\sum_{i=1}^{n} \hat{p}_{i,k}}, \tag{3.5}$$

$$\sigma_{k,d}^{(m+1)} = \frac{\sum_{i=1}^{n} \hat{p}_{i,k} (x_{i,d} - \mu_{k,d}^{(m+1)})^2}{\sum_{i=1}^{n} \hat{p}_{i,k}}, \textit{and} \tag{3.6}$$

$$\boldsymbol{\Sigma}_k^{(m+1)} = \begin{bmatrix} \sigma_{k,1}^{(m+1)} & 0 & \cdots & 0 \\ 0 & \sigma_{k,2}^{(m+1)} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{k,D}^{(m+1)} \end{bmatrix}, \tag{3.7}$$

where $\hat{p}_{i,k} = \begin{cases} 1 & \text{if } y_i^{(m+1)} = k \\ 0 & \text{else} \end{cases}$.

These steps repeat until the algorithm converges, i.e., when no voxel changes classification from one iteration to the next.

3.3.2 Method for Automatic Segmentation

The MRI output in this study consists of seven images of a specific slice of the spinal column—one T2W image and six DWIs from independent diffusion gradient directions. Figures 3.2 and 3.3 show examples of these images for a control animal and an injured animal, respectively. We choose to use the DWIs for all cord/background tissue segmentation because of their useful properties, as described in the Background section.

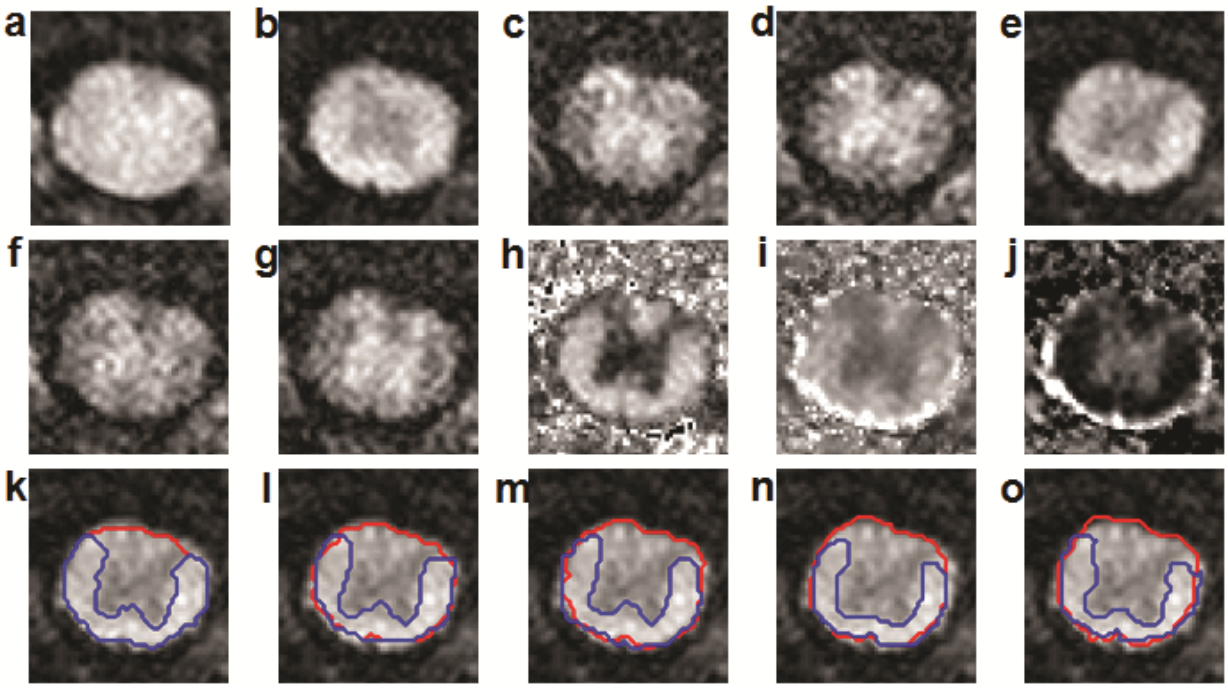

Figure 3.2: Representative example of cropped and normalized MRI data for a control spine slice: (a) T2W image; (b)-(g) diffusion weighted images (DWIs); (h)-(j) DTI maps (relative anisotropy (RA), axial diffusivity ($\lambda_{\parallel}$), and radial diffusivity ($\lambda_{\perp}$), respectively); (k)-(o) manual segmentations of spinal cord (red curves) and spared white matter (blue curves), superimposed over the DWI in (b), for the five separate manual segmentations.

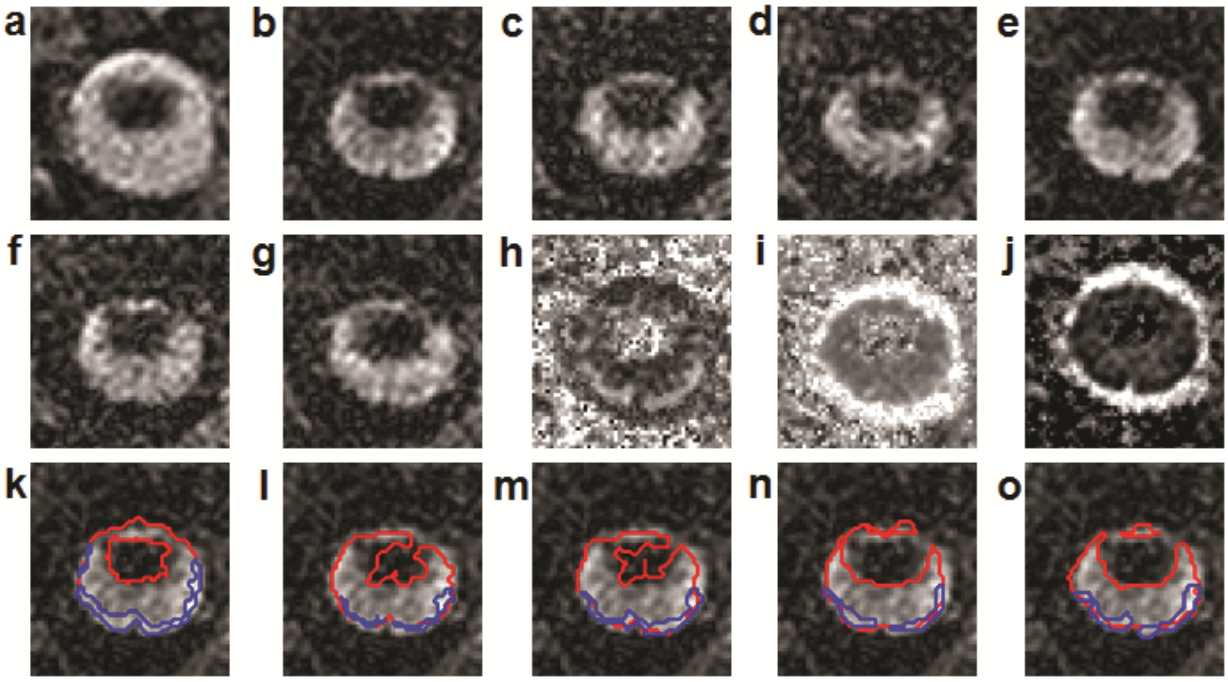

Figure 3.3: Representative example of cropped and normalized MRI data for an injured spine slice: (a) T2W image; (b)-(g) diffusion weighted images (DWIs); (h)-(j) DTI maps (relative anisotropy (RA), axial diffusivity ($\lambda_{\|}$), and radial diffusivity ($\lambda_{\perp}$), respectively); (k)-(o) manual segmentations of spinal cord and hemorrhage (red curves) and spared white matter (blue curves), superimposed over the DWI in (b), for the five separate manual segmentations.

Find initial spinal cord segmentation In the T2W image of each spinal slice (Figs. 3.2.a and 3.3.a), we use the CEM algorithm to classify voxels into two sets, background and spinal cavity. We refer to this set of spinal cavity voxels as $\mathbf{S}_{\text{T2W}}$. This step provides a loose constraint on the spinal cord's location, since the T2W image does not generally differentiate between CSF and spinal cord, so the bright region will be larger than the cord but should completely encapsulate it.

Next, we apply the multidimensional CEM algorithm to the set of DWIs (Figs. 3.2.b-g and 3.3.b-g) to jointly classify voxels as background or as spinal cord according to all six DWIs. We refer to this set of spinal cavity voxels as $\mathbf{S}_{\text{DWI}}$. This step more accurately separates the spinal cord from the background than the T2W image step. By accepting voxels as spinal cord, $\mathbf{S}_0$, only when they are so classified according to both the T2W image

and the DWIs, i.e.,

$$\mathbf{S}_0 = \mathbf{S}_{\text{T2W}} \cap \mathbf{S}_{\text{DWI}}, \tag{3.8}$$

most outlying bright spots are correctly classified as background, without loss of correctly classified cord voxels.

Constrain by location We automatically classify outlying voxels as background by adding a localization constraint. In most cases, the previous two steps will cleanly differentiate between background and spinal cord. This step accounts for potential colocated bright areas in both image types, such as nerve tissues, that are not actually within the cord.

To correctly classify bright, noncord areas as background, we reclassify all small bright areas in $\mathbf{S}_0$ as background, a process demonstrated in Figure 3.4. First, we find the set of boundary curves $\mathbf{B}_0$ that separate the voxels currently classified as spinal cord from the rest of the image (Fig. 3.4.c,i,o). Finding these boundary curves is very simple, unlike the usual contour methods, because we find the boundaries using edge detection on the binary image of voxel classifications, in which voxels have an intensity value of 1 if they are members of $\mathbf{S}_0$ and 0 intensity otherwise. Then, we refine the set of spinal cord voxels by including in the new set $\mathbf{S}_1$ only the voxels in $\mathbf{S}_0$ that lie within the largest such curve ($\mathbf{B}_{0,\text{max}}$) (Fig. 3.4.d,j,p).

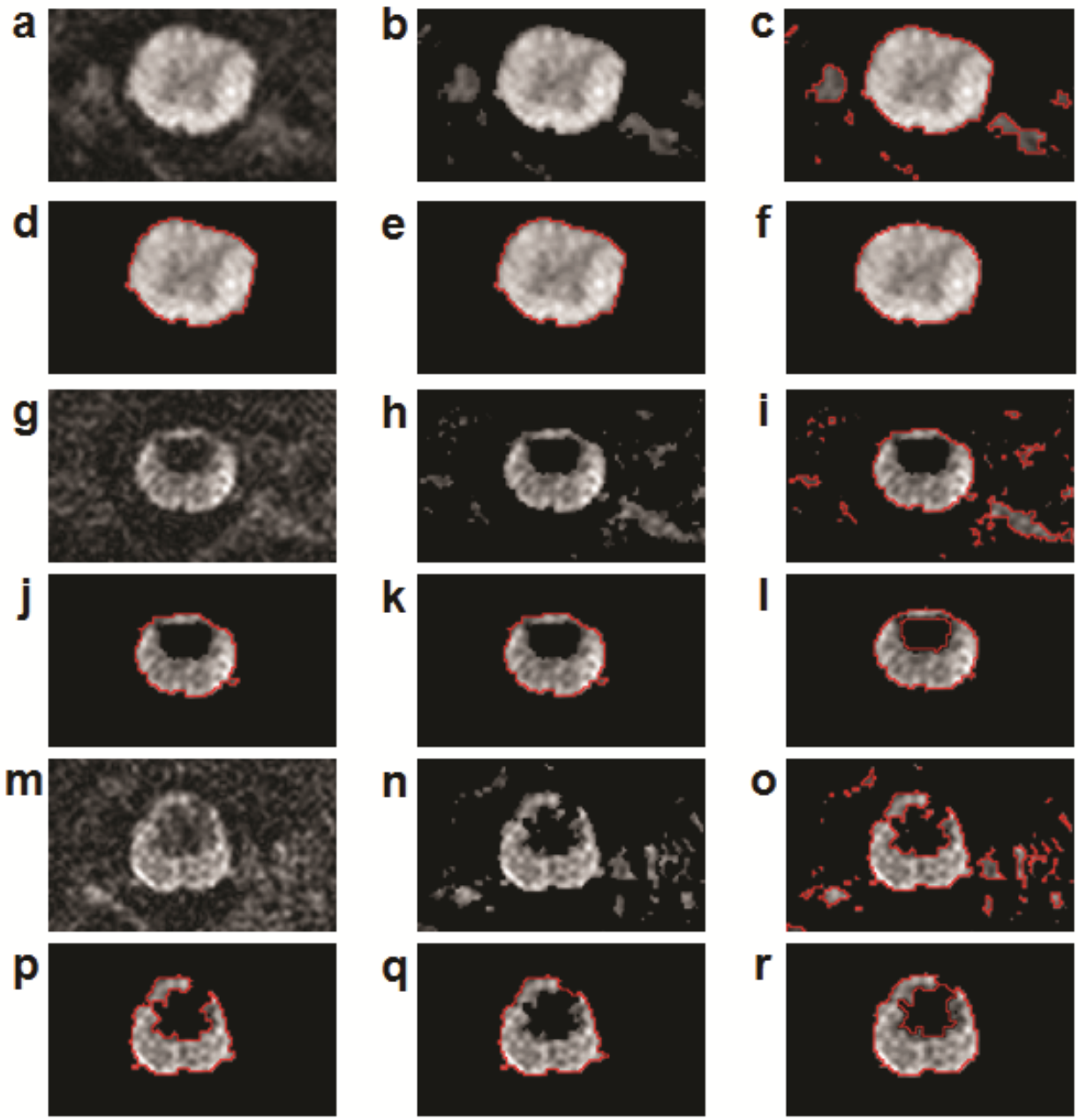

Figure 3.4: Representative examples of spine and hemorrhage segmentation for a control spine slice (a-f) and for injured spine slices with hemorrhage within (g-l) and along (m-r) the spinal cord boundary: (a,g,m) sample cropped, normalized DWI; (b,h,n) after application of the CEM algorithm to the DWIs; (c,i,o) with boundary curves outlined; (d,j,p) after removal of small boundary curves; (e,k,q) after inclusion of holes from T2W image; (f,l,r) after application of composite superellipse bounding shape, with hemorrhage also outlined.

Account for regions of hemorrhage If a region of hemorrhage exists that is completely contained within nonhemorrhaging spinal cord tissue, as in Figure 3.4.g-l, generating two segmentations, one excluding the hemorrhage ($\mathbf{S}_{\text{excl}}$) and one including it ($\mathbf{S}_{\text{incl}}$), does not require additional work. The segmentation excluding hemorrhage is simply

$$\mathbf{S}_{\text{excl}} = \mathbf{S}_1, \tag{3.9}$$

while in this case, the segmentation including hemorrhage is the set of all voxels enclosed by $\mathbf{B}_{0,\max}$.

Unfortunately, regions of hemorrhage can lie along the cord boundary, as in Figure 3.4.m-r, and in such cases, producing the second segmentation is more challenging. However, because the whole cord including hemorrhage is surrounded by CSF, it is possible to identify regions of hemorrhage by locating holes in the set $\mathbf{S}_{\text{T2W}}$, which contains spinal cord and all of its surrounding CSF, even if the hemorrhage is not encapsulated by surviving spinal cord. To identify these holes, first, we find the set of boundary curves $\mathbf{B}_{\text{T2W}}$ around the voxels in $\mathbf{S}_{\text{T2W}}$. Next, we exclude all outliers, retaining only the largest boundary curve $\mathbf{B}_{\text{T2W,max}}$. $\mathbf{B}_{\text{T2W,max}}$ therefore encapsulates spinal cord, hemorrhage, and CSF. We then classify as hemorrhage, $\mathbf{H}$, all points encapsulated by $\mathbf{B}_{\text{T2W,max}}$ that are not members of the set $\mathbf{S}_{\text{T2W}}$, thus removing the voxels containing CSF and surviving spinal cord.

By combining the two sets, that of hemorrhage $\mathbf{H}$ and that of cord excluding hemorrhage $\mathbf{S}_{\text{excl}}$, we are able to find a segmentation for the entire spinal cord,

$$\mathbf{S}_{\text{incl}} = \mathbf{S}_{\text{excl}} \cup \mathbf{H}. \tag{3.10}$$

The final boundary curve around the entire cord is then simply the curve separating the voxels in $\mathbf{S}_{\text{incl}}$ from the rest of the image. This step is applied to all cords. In the case of healthy spine, this step will not change the segmentation. In the case where there are regions of hemorrhage, this step allows us to identify the actual boundary of the spinal cord along with precisely where hemorrhage is, and therefore allows us to calculate, for instance, the amount of hemorrhage relative to the area of the total cord.

Apply a bounding shape The final step in segmenting the spinal cord is the application of a bounding shape. We fit a modified superellipse to the set of voxels in $\mathbf{S}_{\text{incl}}$ to remove possible protrusions from the cord, such as nerve tissues branching away from the spine. This modified superellipse is described as follows:

$$\begin{aligned} \left|\frac{(x-x_0)}{a}\right|^2 + \left|\frac{(y-y_0)}{b}\right|^2 &\leq 1 \quad \text{if } y > 0, \\ \left|\frac{(x-x_0)}{a}\right|^3 + \left|\frac{(y-y_0)}{b}\right|^3 &\leq 1 \quad \text{if } y < 0, \end{aligned} \tag{3.11}$$

where (x_0, y_0) is the center of the superellipse, a is the horizontal semidiameter, and b is the vertical semidiameter. This pair of equations generates a shape that can somewhat tightly encapsulate the spinal cord over the range of shapes it takes in varying states of injury and at varying points along its length. No tight bounding shape can be chosen because of the wide array of injured spinal cord shapes.

Segment spared white matter The final step is the automatic classification of the spared white matter within the segmented cord tissue (Fig. 3.5). We use the multidimensional CEM algorithm to create an initial spared white matter segmentation. Next, we apply a localization constraint by first finding boundary curves around each disjoint group of voxels that the CEM step classified as spared white matter and then reclassifying the smallest such groups into the gray/injured white matter classification. For this step, only voxel groups with a boundary length of at least 20 pixels retain their classification as spared white matter.

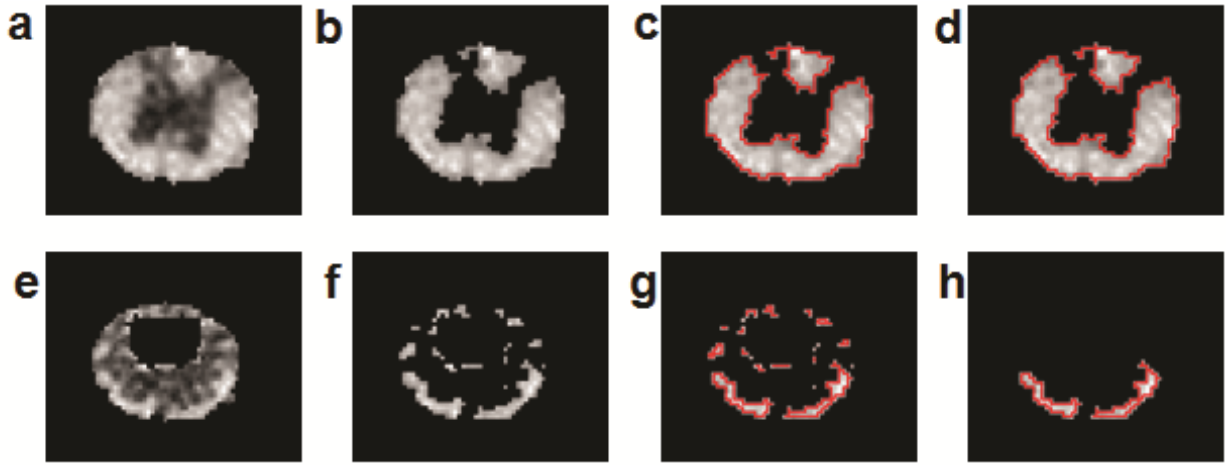

Figure 3.5: Representative examples of spared white matter segmentation for control (a-d) and injured (e-h) spine slices: (a,e) sample cropped, normalized relative anisotropy (RA) map after spinal cord segmentation; (b,f) after application of the CEM algorithm to the three DTI maps; (c,g) with boundary curves outlined; (d,h) after removal of small boundary curves.

3.3.3 Rodent Spinal Cord MRI Experiments

We evaluated the segmentation performance of our algorithm using *in vivo* MR images of uninjured and injured rodent spinal cords.[1,2] Ten twelve-week-old female C57BL/6 mice weighing 18 $\sim$ 20 g (Harlan, Indianapolis, IN) were anesthetized with an isoflurane and oxygen mixture (7% for knock out and 1.5% for maintenance). After dorsal laminectomy at the T8 and T9 vertebral levels, the mice received contusive spinal cord injury utilizing a modified Ohio State University device [27]. The injury group underwent contusion injury at 0.2 m/s with 0.6 mm impact displacement. After impact, the site was closed in layers with 4-0 silk sutures. Enrofloxacin (2.5 mg/kg) and lactated Ringer's (1 ml) were administered subcutaneously. The control group received sham operations including laminectomy and

[1]All surgical interventions and both pre- and post-surgical care were performed in accordance with the Public Health Service Policy on Humane Care and Use of Laboratory Animals, *Guide for the Care and Use of Laboratory Animals* (Institute of Laboratory Animal Resources, National Research Council, 1996), and with the approval of the Washington University Institutional Care and Use Committee.

[2]Standard postoperative care including bladder expression was provided in accordance to the manual of Spinal Cord Injury Research Training Program held at Spinal Trauma and Repair Laboratories in Ohio State University (http://medicine.osu.edu/sci/).

zero point contact of impactor tip on the surface of the spinal cord to establish a reference position, but no impact.

Animal preparation for *in vivo* DTI All mice were delivered to the MR facility and anesthetized with an isoflurane and oxygen mixture (1.0 - 1.5% for maintenance) at 14 days postinjury. The body temperatures were maintained at 37°C with a circulating warm water pad. An inductively coupled surface coil covering T8 - T10 vertebral segments (15 mm × 8 mm) was used as the radio frequency receiver. A 9-cm-inner-diameter Helmholtz coil was employed as the radio frequency transmitter. The entire preparation was placed in an Oxford Instruments (Oxford, UK) 200/330 magnet (4.7 T, 33-cm clear bore) equipped with an actively shielded, Magnex Scientific (Oxford, UK) gradient coil (10–15-cm-inner-diameter, 18 G/cm, 200-μs rise time). The magnet, gradient coil, and Techron gradient power supply were interfaced with a Varian UNITY-INOVA console (PaloAlto, CA) controlled by a Sun Microsystems Blade 1500 workstation.

***In vivo* DTI** A conventional spin-echo imaging sequence was modified by adding Stejskal-Tanner diffusion weighting gradients [56]. The pulse repetition time (∼1.2 sec) was varied according to the period of the respiratory cycle (∼270 ms). The spin echo time = 38 ms, time between application of gradient pulses (Δ) = 20 ms, and diffusion gradient on time (δ) = 7 ms were fixed throughout the experiment. For each animal, three consecutive slices were collected to cover the epicenter of the contusion-injured cord, with a total scan time of 2 hours. DWIs were obtained with diffusion sensitizing gradients applied in six orientations, (G_x,G_y,G_z) = $(1,1,0)$, $(1,0,1)$, $(0,1,1)$, $(-1,1,0)$, $(0,-1,1)$, and $(1,0,-1)$, using diffusion sensitizing factors (b values) of 1.0 ms/μm^2. One image (the b=0 or T2W image) was collected without diffusion sensitizing gradient to serve as a reference. Six scans were averaged per k-space line. The field of view was 10×10mm^2 with 1.0mm slice thickness and the image data matrix for each slice was 128 (phase encoding)× 256 (read out) (zero filled to 256×256). Of the total image area for control animals, on average 1100 voxels contain spinal cord.

A weighted linear least-squares method was used to estimate diffusion tensors for each voxel from the DWIs [34]. The eigenvalue decomposition was then applied to each tensor, yielding a set of eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and eigenvectors for each voxel. Maps of diffusion

indices including relative anisotropy (RA) and axial and radial diffusivities ($\lambda_\parallel$ and $\lambda_\perp$) were generated by applying the following equations for each voxel:

$$\lambda_\parallel = \lambda_1, \tag{3.12}$$

$$\lambda_\perp = \frac{\lambda_2 + \lambda_3}{2}, \tag{3.13}$$

$$< \mathrm{D} > = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}, \tag{3.14}$$

$$\mathrm{RA} = \frac{\sqrt{\sum_{i=1}^{3} (\lambda_i - < \mathrm{D} >)^2}}{\sqrt{3} < \mathrm{D} >}. \tag{3.15}$$

3.3.4 Quantification of Segmentation Accuracy

We automatically segmented the spinal cords and the white matter from the MRI images of all ten mice at each of the spinal slice locations. Because the MRI data we use to test our algorithm have a very large field of view relative to the size of the spine (the area of the spine is roughly 2% of the entire field of view), the data are manually cropped prior to application of the algorithm to a rectangle around the spine. This cropping yields an image for segmentation in which the spine comprises roughly 20% of the field of view.

The spinal cord contains an area of white matter called the dorsal column, which is disconnected from the rest of the white matter and is not included in the white matter segmentations of our experts. Because of this, to compare our automatic segmentations to the manual segmentation, the dorsal column needs to be excluded from the automatic white matter segmentations. To eliminate this area, we automatically exclude the pixels in a small trapezoidal section of the spinal cord, defined relative to the superellipse of (3.11), with four vertices at $(x_0 - a/4, y_0 + b/2)$, $(x_0 + a/4, y_0 + b/2)$, $(x_0 + a/2, y_0 + b)$, and $(x_0 - a/2, y_0 + b)$. We then apply a localization constraint to remove any small sections of dorsal white matter than may not have been fully contained in the trapezoid.

Our algorithm ran in an average of 0.709 seconds (using DWIs for all segmentation) per spinal slice using MATLAB (Mathworks, Natick, MA) on an Intel Core 2 Quad CPU, 2.4 GHz personal computer. We compared the performance of the algorithm when using only DTI maps, using only DWIs, or using both for the spared white matter segmentation. The

algorithm ran in an average of 0.448 seconds when using DWIs for the cord/hemorrhage tissue segmentations and DTI maps for the spared white matter segmentation, and 0.515 seconds when using DWIs for the cord/hemorrhage tissue segmentations and all data (DWIs and DTI maps) for the spared white matter segmentation.

For verification of our algorithm's performance, three types of manual segmentations were created. The entire spinal cord (gray and white matter), the cord excluding hemorrhage, and the white matter were manually segmented by five experts for both control and injured cords utilizing DWIs and calculated diffusion maps. This was done for one spinal slice from the MR images of each animal, at the location of the sham operation for control animals and at the epicenter of surgically induced SCI for injury group animals. From the five expert segmentations for each tissue type, we are able to find a measure of the variation in manual segmentations across experts, to which we can compare the variation between manual and automatic segmentations.

Because our study uses *in vivo* imaging, there is no ground truth available, such as histology. Lacking a ground truth by which to calculate percent error, we instead evaluate our algorithm using the overlap of our automatic segmentation results with manual segmentations of the same data. This overlap is calculated using the binary classification images, i.e., images that take a value of 1 only the in region of cord, of cord excluding hemorrhage, or of white matter, and a value of 0 outside that region. We calculate overlap as

$$\text{Overlap}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cap \mathbf{B}}{\mathbf{A} \cup \mathbf{B}}, \tag{3.16}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the two binary classification images to be compared.

3.4 Results

Table 3.1 shows the average overlap between each pair of independently drawn manual segmentations. This establishes a baseline for variation in manual segmentations to which the variation between the manual and automatic segmentations (Table 3.2) can be compared. It is clear from the poor correspondence of the injured cord segmentations of hemorrhage

and particularly white matter that manual segmentation cannot be treated as the ground truth, as it is in some works.

Table 3.1: Average Percent Overlap between the Five Manual MRI Tissue Segmentations

	Control Group	Injury Group
Entire Spinal Cord	91.47 ± 1.78	90.43 ± 2.45
Spinal Cord Excluding Hemorrhage	91.47 ± 1.78	68.10 ± 12.53
Spared White Matter	80.21 ± 3.28	51.41 ± 9.14

Values are given as the mean ± standard deviation

Table 3.2: Average Percent Overlap between the Automatic Segmentations and the Five Manual MRI Tissue Segmentations

	Control Group	Injury Group
Entire Spinal Cord	91.32 ± 2.39	90.20 ± 2.13
Spinal Cord Excluding Hemorrhage	91.28 ± 2.27	72.18 ± 9.39
Spared White Matter Using DWIs	73.24 ± 6.80	11.87 ± 8.32
Spared White Matter Using DTI Maps	79.65 ± 5.48	54.82 ± 9.88
Spared White Matter Using All Data	77.76 ± 4.93	37.26 ± 16.13

Values are given as the mean ± standard deviation

Table 3.2 shows the total average overlap of our segmentation results with the manual segmentations. For segmentation of the tissue regions of control spinal cords, the performance of our algorithm is on average equivalent to that of the expert manual segmentations. Similarly, for all segmentations of the injured cords—the whole cords, the spinal cord excluding hemorrhage, and the white matter—our automatic segmentations are on average at least as good as the expert manual segmentations. The extreme lack of consistency between expert segmentations of injured white matter makes it impossible for our algorithm to have a high overlap with all experts, but considering the overlap between expert segmentations, the relative performance is strong.

3.5 Discussion

In comparison of our algorithm to the individual manual segmentations, we note that the maximum overlap between our algorithm and any manual segmentation is always greater than the minimum overlap between any two manual segmentations, for all tissue-type segmentations. That is to say, our algorithm performs at least as well as the worst human expert, given that we do not know which is the most accurate segmentation in the group. It is important to note that, as we do not have a ground truth segmentation, it is not possible to say if a particular manual segmentation is superior or inferior to our automatic segmentation. We can only say with certainty that our algorithm has higher consistency with manual segmentations than the consistency between manual segmentations for injured spinal cords.

We found that, although the properties of DWIs are very useful for segmentation of spinal cord and hemorrhage, our results for white matter segmentation using the DTI maps were both qualitatively and quantitatively more accurate for moderately to severely injured cords and also more robust with respect to image quality. In fact, the relative noise level in the DWIs as compared to the DTI maps is such that consideration of all the data, both DWIs and DTI maps, reduces the segmentation performance.

As a point of comparison, in Ellingson et al. [16, 17] and in Younis et al. [61], intensity-based, fuzzy classification methods are used for spine and white/gray matter segmentation. However, the method in Younis et al. [61] relies on the fact that the spinal cord is excised for finding the spinal mask and so cannot be applied to *in vivo* images. Additionally, they do not measure the agreement between manual segmentations and their segmentation, but compare only the intensity statistics of the two regions, so the segmentation accuracy of their method is not validated even for excised spines. Ellingson et al. [16] attempt to validate their method on five uninjured spinal cords, taking a fixed template as the ground truth. The percent overlap between their automatic results and the fixed template ranged from 84.4% to 89.2% for the spinal cord segmentation, and their percent correct classifications for white and gray matter were 67.1% and 86.5%, respectively. The authors [16] then validate only the manual template-alignment step for estimation of the anisotropy statistics of white matter, gray matter, and CSF from test images, not the final segmentation performance in the case of injured cords.

As a final note, if, instead of assuming Gaussian distributions, we treat the intensity distributions as Rician, our algorithm is slower and performs equivalently with respect to overlap with the manual segmentations. The algorithm is slower because there is no analytical expression for the maximum likelihood estimate for Rician parameters, given samples from the distribution, so an iterative estimation method is necessary. This iterative method must be applied at each iteration of the CEM algorithm, which causes slowing of the total algorithm run time. The average run times are 6.047 seconds for using DWIs for all segmentations, 3.412 seconds using DWIs for the cord/hemorrhage tissue segmentations and DTI maps for the spared white matter segmentation, and 3.462 seconds using DWIs for the cord/hemorrhage tissue segmentations and all data (DWIs and DTI maps) for the spared white matter segmentation. Table 3.3 shows the total average overlap of our segmentation results with the manual segmentations.

Table 3.3: Average Overlap between the Automatic Segmentations Assuming Rician Distributions and the Five Manual MRI Tissue Segmentations

	Control Group	Injury Group
Entire Spinal Cord	91.56 ± 2.22	87.31 ± 3.25
Spinal Cord Excluding Hemorrhage	91.58 ± 2.20	69.60 ± 8.58
Spared White Matter Using DWIs	43.31 ± 22.47	3.61 ± 7.58
Spared White Matter Using DTI Maps	80.14 ± 4.70	54.58 ± 9.24
Spared White Matter Using All Data	79.15 ± 4.58	48.26 ± 8.43

Values are given as the mean ± standard deviation

These results are roughly equivalent to the results when the Gaussian assumption is used because the distributions of the white and gray matter are nearly Gaussian. In addition, although a Gaussian is not a good fit for the background distribution, it can be seen in Figure 3.1 that the decision threshold for spinal cord and background will not be affected significantly by assuming a Gaussian rather than a Rician distribution.

3.6 Conclusion

We have proposed a new multistep, CEM-based approach to spinal cord and white matter segmentation from *in vivo* MR images and we have validated that its performance is on par with that of expert manual segmentation. We have demonstrated that our algorithm, unlike previous approaches, remains reliable for spinal cord segmentation in the presence of moderate and severe cord injury, not just extremely mild injury. In addition, we have demonstrated that our algorithm is as reliable as the average human expert for hemorrhage and white matter segmentation for injured rodent spinal cords. Future work includes incorporation of a more detailed physical model for the spinal tissues, as well as comparison of manual and automatic MRI segmentation results to spinal histology results. We expect that our algorithm will yield closer agreement with histology than an average manual expert segmentation does. Future work also includes adaptation of the algorithm to MR images of human spinal cords, which have lower resolution than MR images of rodent cords.

3.7 Acknowledgements

This work was supported in part by the NIH under Grant NS047592, in part by a National Science Foundation (NSF) Graduate Research Fellowship, and also in part by a Mr. and Mrs. Spencer T. Olin Fellowship for Women in Graduate Study. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NSF.

Chapter 4

Automatic Segmentation of Human Glioblastoma MR Images

Access to large collections of magnetic resonance imaging data enables scientifically rigorous study of cancers like glioblastoma multiforme, the most common form of malignant primary brain tumor, but only if the data can be analyzed. In order to process and analyze large sets of MR imaging data, automated segmentation methods are required. In this chapter, I propose an efficient and effective automated segmentation method, the Enhanced Classification Expectation Maximization (ECEM) algorithm. The ECEM algorithm is novel in its ability to introduce spatial information into the classical CEM algorithm with low additional computational complexity. I compare the ECEM's performance on simulated data to the standard finite Gaussian mixture EM algorithm, which is not spatially aware, and to the hidden-Markov random field EM (HMRF-EM) algorithm, a commonly-used spatially aware automated segmentation method for MR brain images. I also show sample results demonstrating the ECEM algorithm's ability to segment MR images of glioblastoma.

4.1 Introduction

Glioblastoma multiforme is the most common malignant primary brain tumor and is almost always fatal—the 10 year survival rate is only 2.3% [15]. Median survival time without aggressive treatment is less than three months, and with aggressive treatment, only fourteen [29]. Because of the disease's prevalence and poor prognosis, a better understanding of glioblastoma is critical.

In order to more fully understand cancer, and thereby to aid in the development of treatments and diagnostic techniques, The National Cancer Institute (NCI) and the National Human Genome Research Institute, both parts of the National Institutes of Health, together released a large collection of anonymized gene data from various cancers, including cerebral glioblastoma, as part of The Cancer Genome Atlas (TCGA) initiative. In conjuction with this, the Cancer Imaging Archive (TCIA), a project funded by the NCI and hosted at Washington University in St Louis, released a collection of Magnetic Resonance (MR) images for the TCGA glioblastoma patient set. Access to a large collection of patient data enables scientifically rigorous study, but only if the data can be analyzed. However, the very size of the TCIA glioblastoma data set makes manual analysis impractical, and bias in manual analysis can make quantitative analysis unreliable. In order to process and analyze data sets like the TCIA gliblastoma data set, fast and accurate automated methods are required.

Accurate quantification of brain tumor is important at all stages of cancer research and treatment—in preclinical studies, for the evaluation of treatment efficacy; during patient care, for diagnosis and treatment monitoring; and in postclinical analysis of data sets such as the TCIA data set. Automated methods can aid researchers in performing this quantitative analysis via segmentation. By segmenting the brain tissue types—e.g., white matter, gray matter, and cerebral-spinal fluid— and also the abnormal regions—e.g., contrast-enhancing tumor, necrotic core, and edema— it is possible, for instance, to quantify a patient's response to the current treatment. This information in turn enables clinicians to make objective, well-informed decisions about how to proceed with future treatment.

FAST is a commonly used toolkit for MRI brain segmentation, from the Oxford Centre for Functional MRI of the Brain's (FRMIB) Software Library (FSL) [28]. It uses a hidden Markov random field (HMRF) model and the expectation-maximization (EM) algorithm [62]. The EM algorithm classifies MR image voxels into classes based on estimated intensity distributions. However, unlike the standard finite mixture EM (FM-EM) algorithm [60], which is spatially agnostic, the use of a hidden Markov random field model incorporates spatial information into the algorithm. The HMRF-EM framework allows for accurate and robust segmentation, even in the presence of noise inherent in MR images. However, segmentation in the HMRF-EM framework requires additional steps at each iteration of the traditional EM-algorithm, and one of those steps is itself iterative. These additional iterations make the computational cost potentially prohibitive for processing large amounts of data.

One of the key features of MR imaging is the availability of different imaging methods that can be used to differentiate tissue based on its structural and molecular properties. For instance, diffusion tensor imaging (DTI) captures information about the diffusion characteristics of tissues—e.g., white matter is highly anisotropic in its diffusivity, while gray matter is more or less isotropic, which translates to a difference in DT image intensity between the two tissue types. The various MR methods allow differentiation with greater or lesser contrast between different tissue types. This feature of MR imaging only exacerbates the problem of large data volumes as each 3D brain scan comprises not only many 2D slices, but multiple channels for each 2D slice, one per imaging method utilized. Thus the volume of data for even a single 3D scan can be very large. Advances in medical imaging technology promise to compound this issue as the volume of imaging data increases, for example, with longitudinal studies and time series data such as that from fMRI. Based on these factors, automated segmentations techniques must strike a balance between computational efficiency and accurate segmentation, even in the presence of noise common to MR images, in order to process the volume of data emerging medical imaging technologies demand.

In this chapter I propose a variant of the Classification EM (CEM) algorithm, the Enhanced CEM algorithm (ECEM) that incorporates spatial data for segmenting tissue types in MRI data. This method has two novel aspects. First, the ECEM algorithm produces segmentations with only small quality loss compared to the HMRF-EM framework but with much less computational complexity. Second, it can take advantage of multi-channel information in segmenting tissue types without sacrificing computational efficiency. Finally, I show that the ECEM algorithm is appropriate for segmentation of abnormal tissue, not only healthy tissue.

In Section 4.2, I describe the TCIA data set and the HMRF-EM algorithm in more detail. In Section 4.3, I introduce the proposed ECEM algorithm. In Sections 4.4 and 4.5, I present and then discuss results for simulated images and for real MR images of glioblastoma.

4.2 Background

4.2.1 TCIA Data Set

The TCIA data set contains a large collection of anonymized multi-channel MR images of cerebral glioblastoma. Glioblastoma is one of approximately twenty types of cancer chosen for inclusion in the TCGA data set. Like all cancer types chosen for inclusion in the data set, glioblastoma meets three criteria: (1) seriousness of prognosis, (2) breadth of public health impact, and (3) availability of quality data for public release.

For each subject included in the TCIA glioglastoma data set, images were gathered for multiple MR imaging subtypes. Some common MRI channels included in the TCIA glioblastoma data set are T_1 pre-Gd , T_1 post-Gd, FLAIR, DTI, FA map, and ADC map. Gadolinium-based contrast agents are useful for demarcating abnormal areas in the brain. Contrast agents can pass into the brain from the blood stream in places where tumor has compromised the blood-brain barrier. Some Gd-based agents accumulate at tumor sites, target markers of angiogenesis, or target the necrotic parts of a tumor [24]. FLAIR, which stands for FLuid Attenuated Inversion Recovery, reduces the affects of cerebrospinal fluid on the image and reduces contrast between white and gray matter, improving the visibility of lesions and edema [55]. DTI (diffusion tensor imaging), from which FA (fractional anisotropy) and ADC (apparent diffusion coefficient) maps can be calculated, is useful for differentiating among tumor types [58].

Challenges for Automatic Segmentation Despite its potential as a data source, there are several difficulties inherent in working with the TCIA glioblastoma data set, which stem from the fact that it is collected from different imaging centers and MRI machines. The data are very heterogenous—images from different subjects may have different resolutions, contrasts, artifacts, and bias fields. The channels included in each image set also may vary, as may the contrast agents used. In addition, the file labeling conventions differ from subject to subject, making identification of the MR image type in each channel, a crucial step for applying automated segmentation methods, a non-trivial task. Finally, the images may require registration to account for movement of the subject's head during the MRI scan.

This registration attempts to translate and rotate the images from each channel so that the signal in a voxel from any one channel will come from the same volume of brain tissue as the signal in the same voxel from each other channel.

Preprocessed TCIA Data To demonstrate the ability of the ECEM algorithm to segment MR images with large abnormal regions, I use a subset of the TCIA glioblastoma data set provided by David Gutman of Emory, MD, PhD, Assistant Professor of Biomedical Informatics at Emory Healthcare. This subset contains only data for cases where all six of the T_1 pre-Gd , T_1 post-Gd, FLAIR, DTI, FA map, and ADC map channels are present. Several preprocessing steps have been used to prepare the data for automatic segmentation. First, the images are labeled uniformly, so that all of the images can be loaded and identified programmatically, rather than selected and labeled manually. Next, the images are registered so that all channels are aligned with the T_1 pre-Gd channel.

4.2.2 FM-EM and HMRF-EM

In image segmentation, the finite Gaussian mixture model assumes that pixel intensities in an image are samples from a distribution that is a weighted sum of Gaussians. Generally, the proportion, mean, and covariance of each Gaussian are estimated using an expectation-maximization (EM) algorithm. This FM-EM algorithm [60] does not allow for any spatial information, although fixed spatial priors for each pixel, e.g., based on an atlas image, are sometimes used.

The HMRF framework incorporates spatial information into image segmentation by modeling pixel-to-pixel influences as a conditional Markov random field distribution [62]. Under this framework, the influence of each pixel on the pixels in its neighborhood must be estimated, in addition to the estimation of the intensity distribution parameters. FM-EM can be understood as a special, degenerate case of HMRF-EM, where there are no other pixels in any pixel's neighborhood, and therefore no influence between pixels.

FM-EM is a simple and computationally efficient method, but the lack of spatial awareness reduces its ability to find accurate segmentations in the presence of noise. In contrast, because it incorporates spatial awareness, HMRF-EM is very robust to noise but has significant, sometimes prohibitive, computational cost [9].

4.3 Method

I propose a modification of the CEM algorithm that incorporates spatial information - the ECEM algorithm. The steps of the standard CEM algorithm for iteration m are as follows [8], assuming Gaussian distributions:

- E-step: Compute the current posterior probabilities $p_{i,k}$ for all data points $i = 1, \ldots, n$ and all clusters $k = 1, \ldots, K$

$$p_{i,k} = \frac{\pi_k^{(m)} \mathbf{f}(\mathbf{x}_i | \boldsymbol{\mu}_k^{(m)}, \boldsymbol{\Sigma}_k^{(m)})}{\sum_{k'=1}^{K} \pi_{k'}^{(m)} \mathbf{f}(\mathbf{x}_i | \boldsymbol{\mu}_{k'}^{(m)}, \boldsymbol{\Sigma}_{k'}^{(m)})}. \tag{4.1}$$

- C-step: Assign each data point $\mathbf{x}_i$ to the cluster with the largest posterior probability.

$$y_i^{(m+1)} = \arg\max_k \, p_{i,k}. \tag{4.2}$$

- M-step: Compute the maximum likelihood estimates for the parameters, $\pi_k^{(m+1)}$, $\boldsymbol{\mu}_k^{(m+1)}$, and $\boldsymbol{\Sigma}_k^{(m+1)}$:

$$\pi_k^{(m+1)} = \frac{\sum_{i=1}^{n} \hat{p}_{i,k}}{n}, \tag{4.3}$$

$$\boldsymbol{\mu}_k^{(m+1)} = \frac{\sum_{i=1}^{n} \hat{p}_{i,k} \mathbf{x}_i}{\sum_{i=1}^{n} \hat{p}_{i,k}}, \tag{4.4}$$

$$\sigma_{k,d}^{(m+1)} = \frac{\sum_{i=1}^{n} \hat{p}_{i,k} (x_{i,d} - \mu_{k,d}^{(m+1)})^2}{\sum_{i=1}^{n} \hat{p}_{i,k}}, \textit{and} \tag{4.5}$$

$$\boldsymbol{\Sigma}_k^{(m+1)} = \begin{bmatrix} \sigma_{k,1}^{(m+1)} & 0 & \cdots & 0 \\ 0 & \sigma_{k,2}^{(m+1)} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{k,D}^{(m+1)} \end{bmatrix}, \tag{4.6}$$

where $\hat{p}_{i,k} = \begin{cases} 1 & \text{if } y_i^{(m+1)} = k \\ 0 & \text{else} \end{cases}$.

These steps repeat until the algorithm converges, i.e., when no voxel changes classification from one iteration to the next.

4.3.1 ECEM

The standard CEM algorithm is not specifically an image segmentation technique—it is a data clustering algorithm, and as such does not consider the spatial interdependencies within images. To incorporate spatial information, I propose expanding the E-step computation of the posterior probability by adding a voxel-specific prior probability, $\pi_{k,i}^{(m)}$, for each class k. In the expanded M-step, each voxel j in the neighborhood $\mathcal{N}_i$ of voxel i contributes a 'vote' on what class k voxel i should be assigned, with the votes being weighted by a function of the vector $\boldsymbol{\nabla}_{ij}$ from i to j. Choosing this weighting function to be a Gaussian with mean $\boldsymbol{\mu}$

and covariance $\boldsymbol{\Sigma}$, the sum $\pi_{k,i}^{(m)}$ of the weighted votes for class k is given by

$$\pi_{k,i}^{(m)} \propto \sum_{\{j|y_j^{(m)}=k, j\epsilon\mathcal{N}_i\}} exp(-\frac{1}{2}(\boldsymbol{\nabla}_{ij}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\nabla}_{ij}-\boldsymbol{\mu})), \tag{4.7}$$

or, equivalently,

$$\pi_{k,i}^{(m)} \propto \sum_{j\epsilon\mathcal{N}_i} \hat{p}_{j,k}^{(m)} exp(-\frac{1}{2}(\boldsymbol{\nabla}_{ij}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\nabla}_{ij}-\boldsymbol{\mu})). \tag{4.8}$$

Thus, with the additional spatial prior in the expanded E-step, (4.1) becomes

$$p_{i,k}^{(m)} \propto \pi_k^{(m)}\pi_{k,i}^{(m)}\mathbf{f}(\mathbf{x}_i|\boldsymbol{\mu}_k^{(m)}, \boldsymbol{\Sigma}_k^{(m)}). \tag{4.9}$$

A voxel j is defined to lie in the neighborhood $\mathcal{N}_i$ of voxel i if the distance $||\boldsymbol{\nabla}_{ij}||_{\mathbf{W}}$ between the two voxels is less than a set threshold δ. Here the distance is defined as the weighted norm

$$||\boldsymbol{\nabla}_{ij}||_{\mathbf{W}} = \boldsymbol{\nabla}_{ij}^T\mathbf{W}\boldsymbol{\nabla}_{ij}, \tag{4.10}$$

where $\mathbf{W}$ is a diagonal matrix whose entries reflect the relative dimensions of the MRI voxels. Generally, MRI tissue voxels are not cubes—slice depth is often larger than the height and width of the voxel. In a 2D image, $||\boldsymbol{\nabla}_{ij}||_{\mathbf{W}}$ reduces to the Euclidean norm, because the voxels correspond to square regions of tissue.

Similarly, for 2D multi-channel MR images, $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ reduce to $\sigma\boldsymbol{I}$ and $\mu\mathbf{1}$. For 3D or even 4D (3D time-series) data, the relative spatial and temporal distance between voxels in each dimension will be reflected in $\mathbf{W}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\mu}$.

Note that $\pi_{k,i}^{(m)}$ can be very efficiently calculated by convolution of the chosen distance function, here a Gaussian filter, with the binary classification image, where the intensity for voxel i is equal to $\hat{p}_{i,k}^{(m)}$. Therefore, the additional computational cost of the ECEM algorithm relative to the CEM algorithm is K convolutions per iteration. Note also that the proposed ECEM algorithm, like the CEM algorithm, is highly parallelizable. In each iteration, the E-step and C-step for each voxel are completely independent, can can therefore be calculated

in parallel. Thus, this simple proposed extension of the standard, spatially agnostic CEM algorithm integrates spatial awareness in a straightforward and computationally efficient manner.

4.3.2 ECEM Segmentation of Glioblastoma

For the specific case of segmentation of glioblastoma from MR images of human brain, I follow a similar procedure as in Chapter 3.3, but the necessity to remove outlying misclassified voxels post-ECEM is greatly reduced relative to simple CEM. As a preprocessing step, I first automatically isolate the brain from the skull and other background, by applying the ECEM algorithm assuming two classes. I next assume the intensity values in the multi-channel image come from a sum of three Gaussian distributions—those of the white matter, the gray matter, and the abnormal region. Once the abnormal region has been isolated, I then segment the abnormal region. In the image sets tested, segmentation into four classes seemed to best capture the subregions of abnormal tissue.

For the image sets tested, no significant bias field was present. For MRI segmentation where the bias field is significant, a bias-field estimation step can be added to the ECEM algorithm in the same way it is added in HMRF-EM [62].

4.4 Experiments

In order to demonstrate the ECEM algorithm's segmentation capabilities in the presence of noise, I compare segmentation accuracy for the ECEM, the standard FM-EM, the standard CEM, and the HMRF-EM algorithms on simulated data. I show the effect of noise on pixel misclassification rates for three- and five-class simulated images. Next, I show sample results demonstrating the ECEM algorithm's ability to segment MR images of glioblastoma multiforme on real patient multichannel data from the TCIA glioblastoma data set.

4.4.1 Segmentation of Simulated Data

I evaluate the ECEM algorithm's performance on simulated images in comparison to the FM-EM, the CEM, and the HMRF-EM algorithms, using the simulated piecewise-constant images from Zhang, et al. [62].

Three-class Image The first simulated image used for testing is shown in Figure 4.1.a. The pixels in the noiseless image have three intensity values—30, 125, and 220—with proportions of 0.372, 0.299, and 0.329, respectively. Performance is tested in the presence of additive Gaussian noise, with standard deviations of $\sigma = 28$ (CNR=3.4), $\sigma = 47$ (CNR=2.0), and $\sigma = 95$ (CNR=1.0) (Fig. 4.1.b-d). Here contrast-to-noise ratio (CNR) is defined as the average intensity difference from one class mean to the next, divided by the standard deviation σ of the added Gaussian noise. After the addition of Gaussian noise, values that drop below zero or rise above 255 are truncated to zero and 255, respectively.

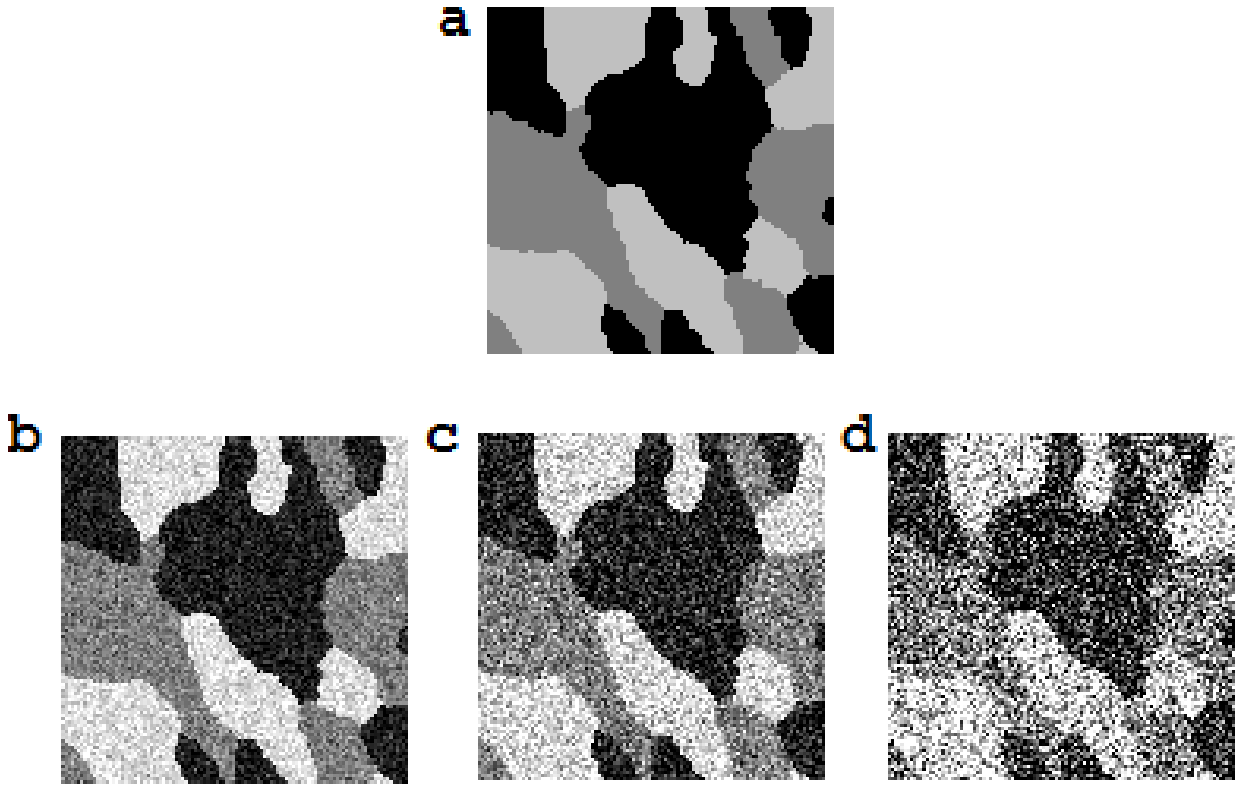

Figure 4.1: Test three-class image. (a) Original, noiseless image. (b)-(d) Image with additive Gaussian noise, CNR = 3.4, 2.0, and 1.0, respectively.

Standard FM-EM converges only for the lowest noise level, with misclassification rate (MCR) of 10.50%, as shown in Figure 4.2 and Table 4.1. The misclassification rate (MCR) is defined simply as the percentage of incorrectly classified voxels.

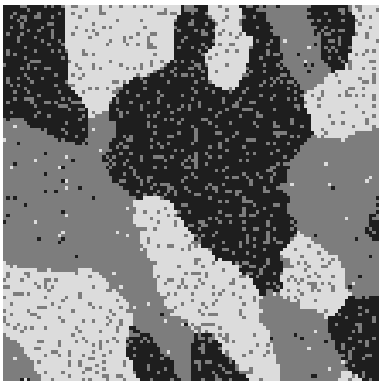

Figure 4.2: Three-class segmentation results from FM-EM for CNR = 3.4, MCR = 10.50%.

Table 4.1: Three-Class Parameter Estimation Using the FM-EM Algorithm

class	Class 1			Class 2			Class 3		
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3
$\sigma = 28$	24.5	19.3	0.301	122.3	43.9	0.422	224.6	21.7	0.277

Standard CEM converges for all three levels, with MCR=5.9%, 20.73%, 39.41%, as shown in Figure 4.3 and Table 4.2.

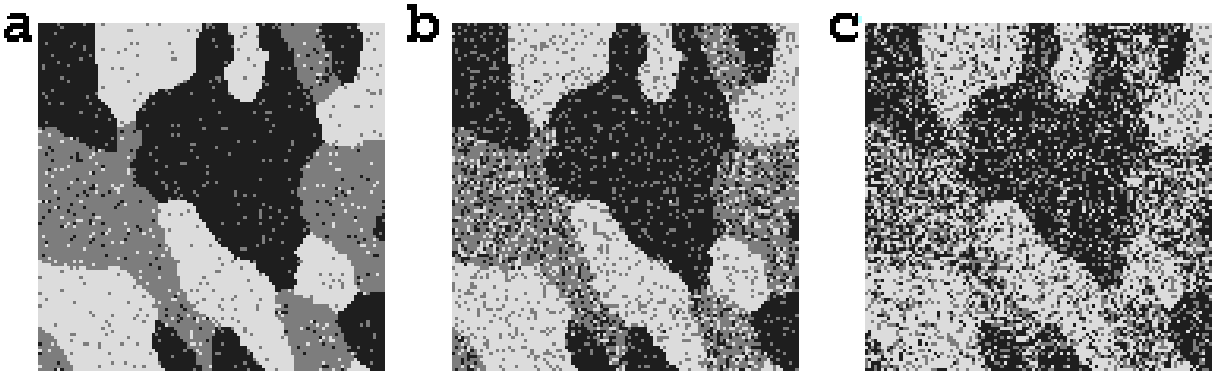

Figure 4.3: Three-class segmentation results from CEM for CNR = (a) 3.4, (b) 2.0, and (c) 1.0. MCR = 5.90%, 20.73%, and 39.41%, respectively.

Table 4.2: Three-Class Parameter Estimation Using the CEM Algorithm

class	Class 1			Class 2			Class 3		
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3
$\sigma = 0$	30.0	0	0.372	125.0	0.0	0.299	220.0	0.0	0.329
$\sigma = 28$	30.0	22.2	0.368	124.6	25.0	0.302	219.9	23.4	0.324
$\sigma = 47$	33.0	28.5	0.392	128.4	24.0	0.279	219.9	28.2	0.329
$\sigma = 95$	27.0	29.9	0.404	126.0	22.5	0.233	227.2	30.3	0.363

HMRF-EM converges for all three levels, with MCR=0.12%, 1.04%, and 8.73%, respectively, as shown in Figure 4.4 and Table 4.3 (reproduced from Zhang, et al. [62]).

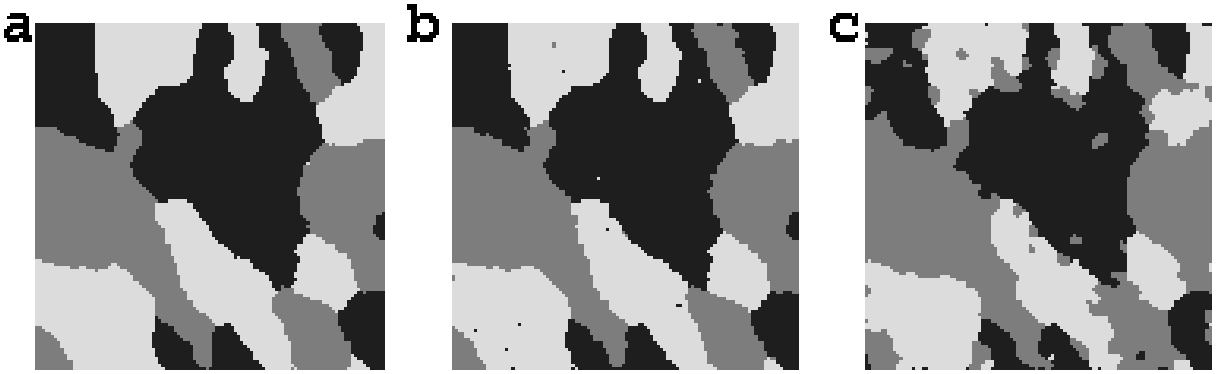

Figure 4.4: Three-class segmentation results from HMRF-EM for CNR = (a) 3.4, (b) 2.0, and (c) 1.0. MCR = 0.12%, 1.04%, and 8.73%, respectively. (reproduced from Zhang, et al. [62])

Table 4.3: Three-Class Parameter Estimation Using the HMRF-EM Algorithm

class	Class 1			Class 2			Class 3		
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3
$\sigma = 28$	32.0	24.6	0.378	124.9	27.8	0.300	219.2	24.8	0.332
$\sigma = 47$	36.1	35.6	0.377	124.6	46.1	0.304	213.5	38.1	0.320
$\sigma = 95$	52.1	61.7	0.355	127.7	80.5	0.363	203.6	62.1	0.281

The ECEM method also converges for all three levels, with MCR=0.40%, 4.07%, and 43.66%, as shown in Figure 4.5 and Table 4.4, although its classification accuracy is nearly identical to that of standard CEM for the lowest CNR case.

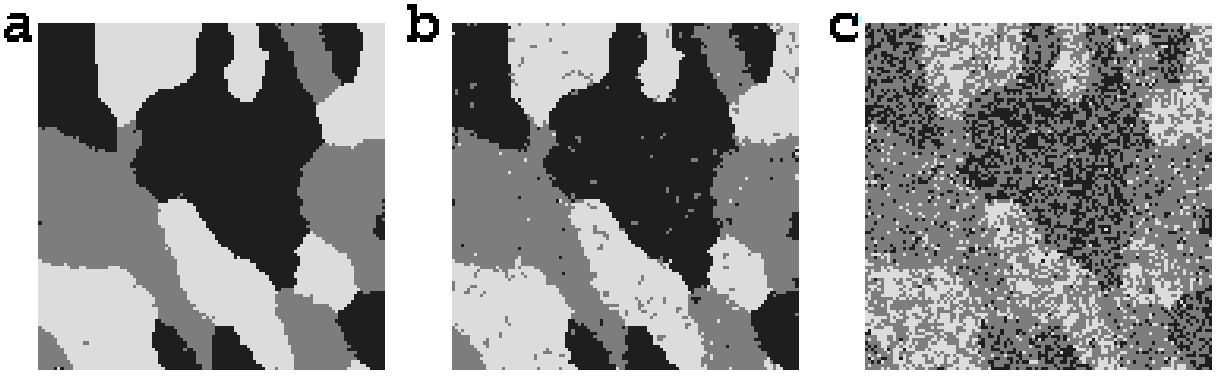

Figure 4.5: Three-class segmentation results from ECEM for CNR = (a) 3.4, (b) 2.0, and (c) 1.0. MCR = 0.40%, 4.07%, and 43.66%, respectively.

Table 4.4: Three-Class Parameter Estimation Using the ECEM Algorithm

class	Class 1			Class 2			Class 3		
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3
$\sigma = 0$	30.0	0	0.372	125.0	0.0	0.299	220.0	0.0	0.329
$\sigma = 28$	31.4	24.3	0.373	125.2	28.2	0.298	219.1	25.1	0.330
$\sigma = 47$	35.5	35.0	0.372	127.2	43.1	0.315	216.6	34.7	0.313
$\sigma = 95$	2.6	9.9	0.199	121.5	63.4	0.604	248.1	18.9	0.197

Five-class Image The second simulated image used for testing (Fig. 4.6.a) is also from Zhang, et al. [62]. The pixels in the noiseless image take on five intensity values—30, 77, 125, 172, and 220—with proportions of 0.280, 0.273, 0.113, 0.187, and 0.147, respectively. Performance is tested in the presence of additive Gaussian noise, with standard deviations of $\sigma = 23$ (CNR=2.0), $\sigma = 33$ (CNR=1.4), and $\sigma = 47$ (CNR=1.0) (Fig. 4.6.b-d). After the addition of Gaussian noise, values that drop below zero or rise above 255 are truncated to zero and 255, respectively.

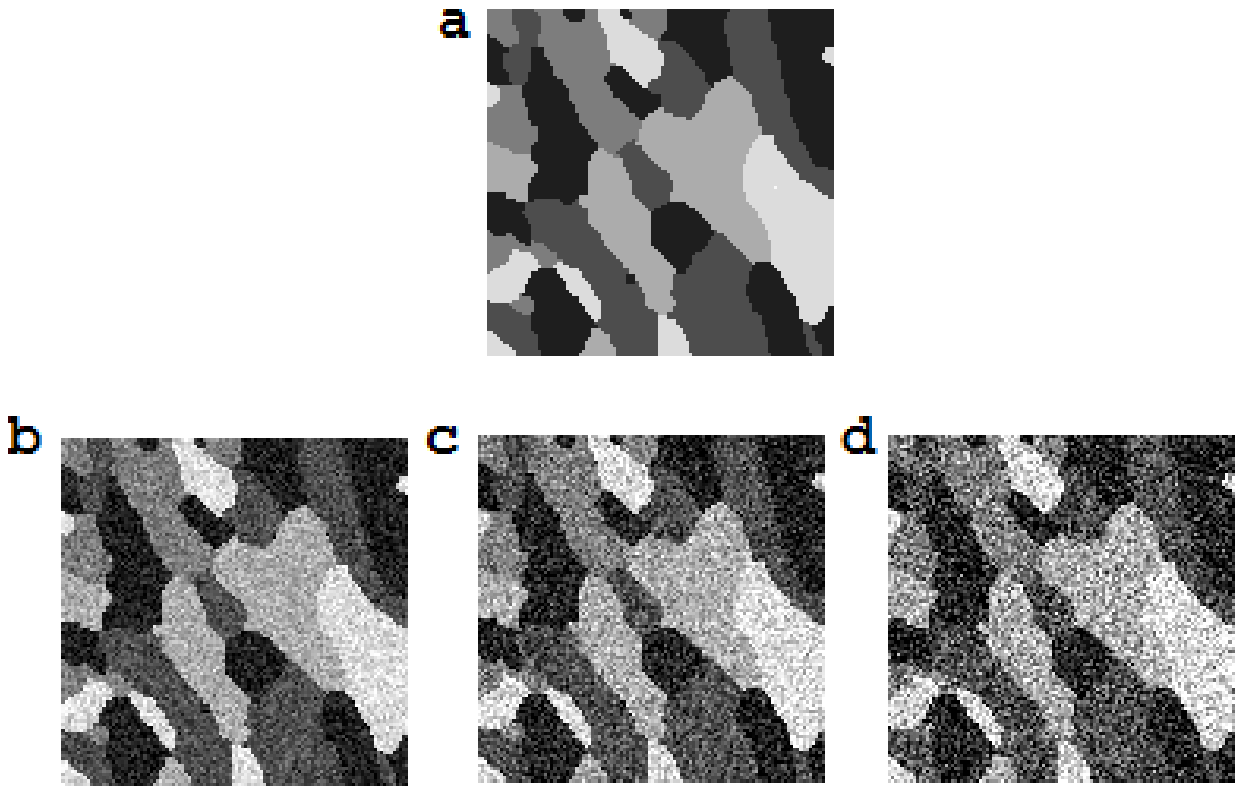

Figure 4.6: Test five-class image. (a) Original, noiseless image. (b)-(d) Image with additive Gaussian noise, CNR = 2.0, 1.4, and 1.0, respectively.

Standard FM-EM does not converge, even for the highest CNR. The CEM method converges for all three levels, with MCR=25.38%, 37.69%, and 48.40%, as shown in Figure 4.7 and Table 4.5.

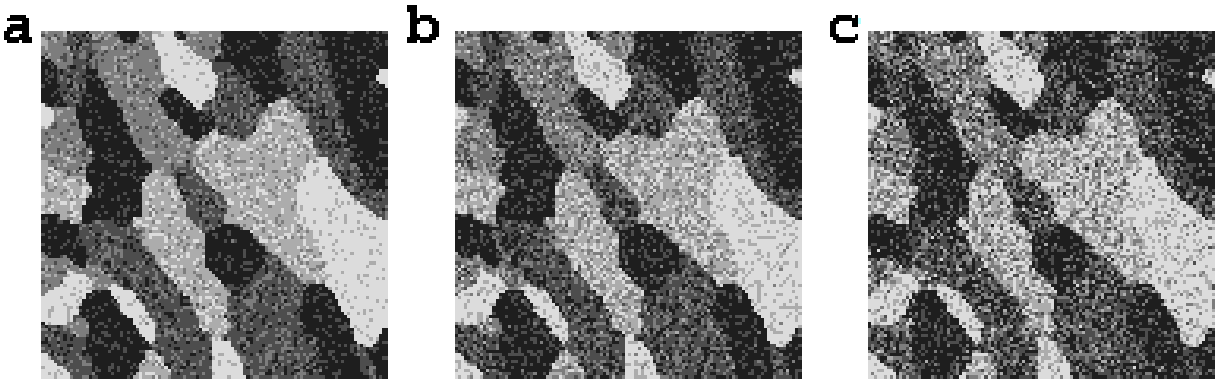

Figure 4.7: Five-class segmentation results from CEM for CNR = (a) 2.0, (b) 1.4, and (c) 1.0. MCR = 25.38%, 37.69%, and 48.40%, respectively.

Table 4.5: Five-Class Parameter Estimation Using the CEM Algorithm

class	Class 1			Class 2			Class 3			Class 4			Class 5		
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3	μ_4	σ_4	ω_4	μ_5	σ_5	ω_5
$\sigma = 0$	30	0	0.28	77	0	0.27	125	0	0.11	172	0	0.19	220	0	0.15
$\sigma = 23$	27	16	0.27	73	12	0.23	119	15	0.17	170	13	0.16	220	18	0.17
$\sigma = 33$	25	18	0.28	77	13	0.23	125	15	0.18	173	13	0.15	226	19	0.16
$\sigma = 47$	21	19	0.29	78	13	0.18	125	14	0.18	171	13	0.14	228	21	0.18

HMRF-EM converges for all three levels, with MCR=0.2%, 1.36%, and 7.68%, as shown in Figure 4.8 and Table 4.6 (reproduced from Zhang, et al. [62]).

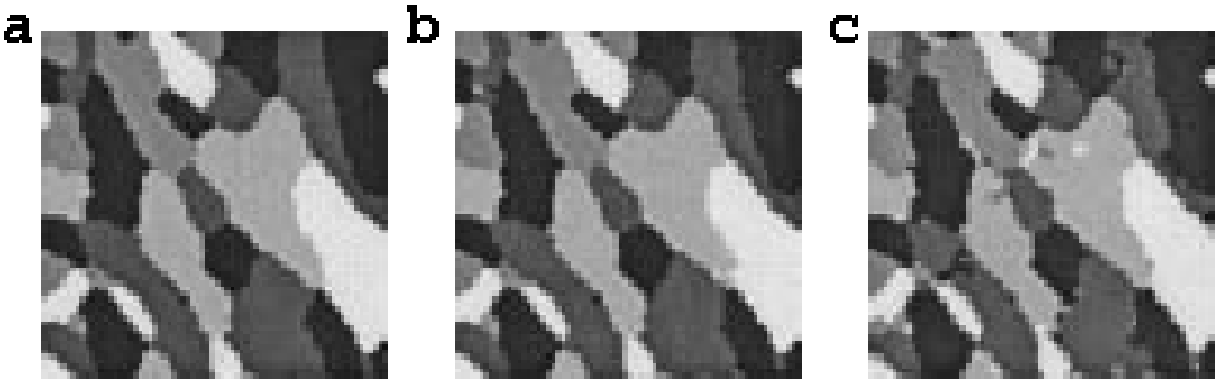

Figure 4.8: Five-class segmentation results from HMRF-EM for CNR = (a) 2.0, (b) 1.4, and (c) 1.0. MCR = 0.2%, 1.36%, and 7.68%, respectively. (reproduced from Zhang, et al. [62])

Table 4.6: Five-Class Parameter Estimation Using the HMRF-EM Algorithm

class	Class 1			Class 2			Class 3			Class 4			Class 5		
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3	μ_4	σ_4	ω_4	μ_5	σ_5	ω_5
$\sigma = 23$	31	22	0.28	76	23	0.27	124	23.5	0.11	171	24	0.19	219	22	0.14
$\sigma = 33$	33	38	0.29	77	32.5	0.26	123	32.6	0.11	171	33	0.19	217	28.5	0.15
$\sigma = 47$	40	37	0.33	81	44	0.25	125	44	0.08	174	45	0.21	216	37	0.12

The ECEM method also converges for all three levels, with MCR=2.26%, 21.93%, and 46.28%, as shown in Figure 4.9 and Table 4.7, although it performs poorly for the lower CNR cases.

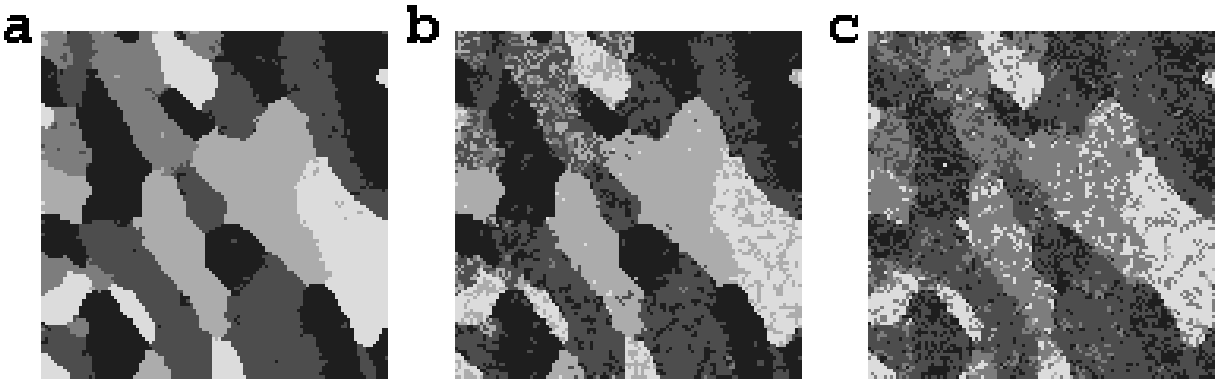

Figure 4.9: Five-class segmentation results from ECEM for CNR = (a) 2.0, (b) 1.4, and (c) 1.0. MCR = 2.26%, 21.93%, and 46.28%, respectively.

Table 4.7: Five-Class Parameter Estimation Using the ECEM Algorithm

class	Class 1			Class 2			Class 3			Class 4			Class 5		
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3	μ_4	σ_4	ω_4	μ_5	σ_5	ω_5
$\sigma = 0$	30	0	0.28	77	0	0.27	125	0	0.11	172	0	0.19	220	0	0.15
$\sigma = 23$	31	20	0.28	78	22	0.27	125	22	0.11	172	23	0.19	220	21	0.15
$\sigma = 33$	33	25	0.33	92	26	0.29	140	9	0.02	174	29	0.26	233	19	0.10
$\sigma = 47$	7	12	0.15	72	34	0.42	157	34	0.28	190	3	0.01	229	26	0.14

4.4.2 Segmentation of MR Images of Human Brains with Glioblastoma

Figure 4.10 illustrates the performance of the ECEM algorithm on multichannel MRI data for glioblastoma. Figure 4.10.a.i-iii shows the input, from three separate patients, to the ECEM segmentation algorithm, the multichannel MR data for a single brain slice. The MR image types used for the segmentations were T_1 (without contrast agent), T_1 post-Gd contrast enhancement (pre-contrast subtracted from postcontrast T_1), FLAIR, DTI, FA map, and ADC map. As described in Section 4.2.1, these images have been registered to the T_1 pre-Gd image. Figure 4.10.b.i-iii shows the results of the ECEM preprocessing to isolate first the

brain (highlighted in teal) from the background, and then the abnormal region (highlighted in red) from the rest of the brain. Figure 4.10.c.i-iii shows the results of four-class ECEM segmentation of the abnormal regions.

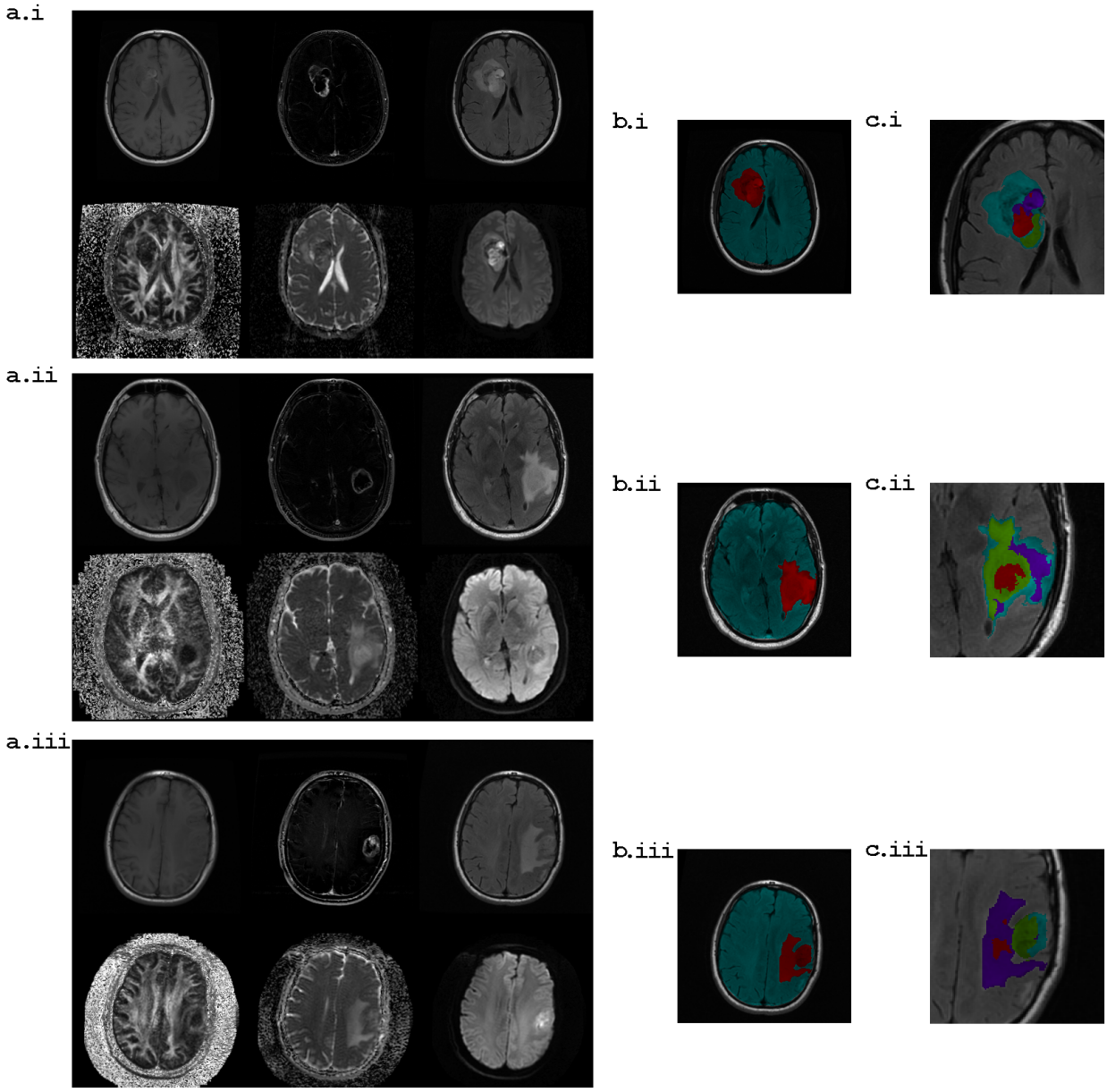

Figure 4.10: (a.i-iii) Multichannel MRI data for a single slice of brain from three subjects. Top row, left to right: T_1 pre-Gd, T_1 post-Gd contrast enhancement (pre-post contrast difference), FLAIR. Bottom row, left to right: DTI, FA map, ADC map. (b.i-iii) Automatic segmentation of whole brain—red is abnormal, teal is normal—overlaid on the FLAIR image. (c.i-iii) Automatic segmentation of abnormal region into four unlabeled tissue classes, overlaid on the FLAIR image.

4.5 Discussion

4.5.1 Segmentation of Simulated Data

Clearly, in comparison with FM-EM, the ECEM algorithm's performance is a drastic improvement at all noise levels. The ECEM algorithm converges where the FM-EM does not. The classification results are significantly more accurate for the one case where FM-EM did converge. The parameter estimation of the ECEM algorithm also outperformed FM-EM.

In comparison with the CEM algorithm, the ECEM enhances classification significantly for the higher CNR cases. The performance gains decrease as noise increases, and in low CNR cases, the misclassification rate of ECEM is on par with CEM. Interestingly, despite the higher misclassification rate, the CEM algorithm actually does a better job of estimating the class means. It is possible that this discrepancy is a byproduct of the way the simulated images were created. In the highest noise images, the distribution of pixel intensities is nearly uniform, and therefore, with no spatial information, the CEM more-or-less uniformly distributes the pixel class means. In these test images, it happens that the actual class means are uniformly distributed, but that is clearly not always the case in real images.

In comparison to HMRF-EM, for low noise levels, the difference in performance is insignificant. As seen in Tables 4.4 and 4.7, the results for the ECEM algorithm and HMRF-EM algorithm are similar in the cases where the CNR is 3.4 and 2.0. For these higher CNR tests, the parameter estimation of the ECEM algorithm was also on par with results shown for the HMRF-EM algorithm in Zhang, et al. [62].

For the three-class image with a CNR of 1.0, the performance of the ECEM algorithm relative to the HMRF-EM algorithm deteriorates sharply. For the five-class case, the inflection point where performance diverges is somewhere between the CNR levels of 2.0 and 1.4. While this shows that there are CNR levels for which HMRF-EM is clearly a better choice, despite its additional computational complexity, a CNR of 1.0 is outside of the range of CNR levels that would realistically be present in brain MR images. This is because, to a certain extent, CNR is controllable in MR imaging—for example, CNR can be increased by increasing the voxel size (i.e., decreasing the image resolution) or by increasing the time spent on the scan [48]. Most importantly, an MR image that is 50% data and 50% noise would be unacceptable for a

clinician or researcher attempting to label tissues or make a diagnosis. While results at this noise level are useful to show the limit to the ECEM approach, the performance degradation at reasonable noise levels (i.e., those that would be present in images for diagnostic or pre-clinical purposes) is very satisfactory. The ECEM algorithm, in exchange for potentially mislabeling a small percentage of voxels compared to HMRF-EM, allows very large data sets to be analyzed more quickly due to its lower computational complexity.

4.5.2 Segmentation of MR Images

The results shown in Figure 4.10 demonstrate the ECEM algorithm's potential for MR brain segmentation of tumor and related abnormalities. Qualitatively, the algorithm appears to correctly differentiate some of the abnormal regions, for example, edema and tumor. However, some regions of interest in the brain images—e.g., the 'ring' of contrast-enhancing tumor which can be seen clearly in the post-Gd contrast images in Figure 4.10.a.i-iii—are not visible in the automatic segmentations. This is at least partially due to the insufficiently precise registration of the image sets. The contrast-enhancing regions are often only two to three voxels wide, and from comparison of set points in the brain between the six channels, the set points frequently shift four or five voxels from channel to channel.

4.6 Conclusion

In this chapter, I have proposed a novel extension to the CEM algorithm that is spatially aware: the ECEM algorithm. The ECEM algorithm incorporates spatial awareness in an elegant, mathematically simple, and computationally efficient manner while producing segmentations that are robust to noise. I compared the performance of the ECEM algorithm to that of the FM-EM, standard CEM, and HMRF-EM algorithms. The experiments presented in this chapter show that the segmentation results from ECEM are significantly better than those from FM-EM, and that incorporating spatial information into the CEM algorithm in general improves the accuracy of the CEM segmentation. For data with high to moderate CNR, the ECEM is a good choice for rapid and robust segmentation, while for very low CNR data HMRF-EM is preferable.

The significance of a robust and efficient segmentation algorithm like the ECEM algorithm is that it allows large data sets to be processed, whether for postclinical analysis of collections such as the TCIA glioblastoma data set, or for clinical use for patient data with multichannel and/or time series components. Fast and efficient methods also may facilitate real-time user interaction, so, for instance, clinicians could manually refine and reseed automatic segmentations. Such an interactive algorithm could combine human expertise with fast and unbiased statistical methods.

Numerous potential modifications of the ECEM algorithm may improve accuracy in high noise data sets with minimal additional computational complexity. For instance, it is possible for the parameters of the Gaussian filter from (4.8) to vary with iteration number; e.g., the neighborhood-based prior can consider a larger neighborhood for early iterations of the ECEM algorithm and a smaller neighborhood for later iterations, or vice versa. Informal testing has demonstrated improvement in final segmentation accuracy when the filter parameters are allowed to vary. Finding an algorithmic method for varying the parameters, rather than manual parameter tuning, remains as potential future work.

4.7 Acknowledgements

This work was supported by a Mr. and Mrs. Spencer T. Olin Fellowship for Women in Graduate Study and NSF Grant CCF-0963742.

Chapter 5

Conclusion

Automatic segmentation techniques based on robust statistical methods are necessary to overcome the challenges of noisy, multichannel MRI data. In this dissertation, I have proposed automated segmentation methods that fall into two classes—segmentation via optimization of a parametric boundary (Chapter 2) and segmentation via multistep, spatially constrained intensity classification (Chapters 3 and 4). These two approaches are applicable in different segmentation scenarios. Parametric boundary segmentation is useful and necessary for segmentation of noisy images where the tissue of interest has predictable shape but poor boundary delineation, as in the case of lung with heavy or diffuse tumor. Spatially constrained intensity classification is appropriate for segmentation of noisy images with moderate contrast between tissue regions, where the areas of interest have unpredictable shapes, as is the case in spinal injury and brain tumor. The proposed automated segmentation techniques address the need for MR image analysis in three specific applications: (1) preclinical rodent studies of primary and metastatic lung cancer, (2) preclinical rodent studies of spinal cord lesion, and (3) postclinical analysis of human brain cancer.

In Chapter 2, I demonstrated that corrected average MR image intensity in mouse lung is an accurate metric of total tumor burden. The tumor measurements were validated by correlating MR image intensities with the weight of the excised lungs. By measuring average MR lung intensity, tumor burden can be estimated *in vivo*, even in cases of diffuse disease where individual tumors cannot be segmented from the MR images. Thus, *relative* measures of tumor burden for a single animal can be established simply by comparing average lung intensities from images collected at different time points. *Absolute* tumor burden measures can also be determined following establishment of a calibration curve between MR image

intensities and lung weights. Because this average image intensity approach requires accurate lung segmentation, I proposed a novel method for automated segmentation of the MR images of murine lungs and pulmonary tumors. I developed a new, two-dimensional parametric model for mouse lung that accurately preserves the overall shape of the lungs, and a novel cost function for optimization of the model parameter values for each lung image. Qualitatively, my segmentation results are well fitted to the lungs. Quantitatively, the correlation between the corrected average intensity tumor-burden metric and the lung weight is excellent, and comparable to that of fully-manual expert segmentation. These promising findings suggest multiple applications of this method. For example, to assess treatment efficacy, clinicians can automatically measure tumor burden over time.

In Chapter 3, I proposed a new multistep, CEM-based approach to spinal cord and white matter segmentation from *in vivo* MR images, and I validated that its performance is on par with that of expert manual segmentation. I demonstrated that my algorithm, unlike previous approaches, remains reliable for spinal cord segmentation in the presence of moderate and severe cord injury, not just extremely mild injury. In addition, I demonstrated that the algorithm is as reliable as the average human expert for hemorrhage and white matter segmentation for injured rodent spinal cords. By replacing manual segmentation, this method will, for example, allow researchers to analyze time-series images of mice to evaluate experimental treatments.

In Chapter 4, I proposed a novel extension to the CEM algorithm that is spatially aware: the ECEM algorithm. The ECEM algorithm incorporates spatial data in an elegant, mathematically simple, and computationally efficient manner while producing segmentations that are robust to noise. I compared the performance of the ECEM algorithm to that of the FM-EM, CEM, and HMRF-EM algorithms. The experiments presented in this chapter show that the segmentation results from ECEM are significantly better than those from FM-EM, and that incorporating spatial information into the CEM algorithm in general improves the accuracy of the CEM segmentation. For data with high to moderate CNR, the ECEM is a good choice for rapid and robust segmentation, while for very low CNR data HMRF-EM is preferable. The ECEM method will allow fast analysis of very large data sets, such as MRI collections which may contain time-series, high resolution, multichannel image sets.

In this dissertation, I identified several potential areas for future work. The lung segmentation method presented requires manual input in order to seed segmentation; fully automating the segmentation would speed analysis. In addition, the lung segmentation method could be successfully adapted to other problems where parametric models are of use, including prostate cancer analysis. My spinal tissue segmentation methods could be improved by incorporation of a more detailed physical model for the spinal tissues, and comparison of manual and automatic MRI segmentation results to spinal histology results could validate these segmentations relative to ground truth. In addition, adaptation of this algorithm to MR images of human spinal cords, which have lower resolution than MR images of rodent cords, is another potential research area. The proposed ECEM algorithm also presents several future research directions. Informal testing showed that allowing algorithm parameters to vary between iterations of the ECEM algorithm improved results. Exploration of these parameters is left as future work.

References

[1] D. M. Basso. Neuroanatomical substrates of functional recovery after experimental spinal cord injury: Implications of basic science research for human spinal cord injury. *Phys. Ther.*, 80(8):808–817, Aug 2000.

[2] D. M. Basso, M. S. Beattie, and J. C. Bresnahan. Descending systems contributing to locomotor recovery after mild or moderate spinal cord injury in rats: Experimental evidence and a review of literature. *Restor. Neurol. and Neurosci.*, 20(5):189–218, 2002.

[3] D.M. Basso, M.S. Beattie, and J.C. Bresnahan. Graded histological and locomotor outcomes after spinal cord contusion using the NYU weight-drop device versus transection. *Exp. Neurol.*, 139(2):244–256, Jun 1996.

[4] Thomas Böttger, Tobias Kunert, Hans P. Meinzer, and Ivo Wolf. Application of a new segmentation tool based on interactive simplex meshes to cardiac images and pulmonary MRI data. *Acad. Radiol.*, 14(3):319–329, 2007.

[5] Matthew D. Budde, Joong Hee Kim, Hsiao-Fang Liang, John H. Russell, Anne H. Cross, and Sheng-Kwei Song. Axonal injury detected by *in vivo* diffusion tensor imaging correlates with neurological disability in a mouse model of multiple sclerosis. *NMR Biomed.*, 21(6):589–597, 2007.

[6] Jack D. Bui, Ravindra Uppaluri, Chyi-Song Hsieh, and Robert D. Schreiber. Comparative analysis of regulatory and effector T cells in progressively growing versus rejecting tumors of similar origins. *Cancer Res.*, 66:7301–7309, 2006.

[7] Beatriz M. Carreno, Joel R. Garbow, Grant R. Kolar, Erin N. Jackson, John A. Engelbach, Michelle Becker-Hapak, Leonidas N. Carayannopoulos, David Piwnica-Worms, and Gerald P. Linette. Immunodeficient mouse strains display marked variability in growth of human melanoma lung metastases. *Clin. Cancer Res.*, 15(10):3277–3286, 2009.

[8] Gilles Celeux and Gerard Govaert. A classification EM algorithm for clustering and two stochastic versions. *Comput. Stat. Data Anal.*, 14(3):315–332, Oct 1992.

[9] Terrence Chen and Thomas S. Huang. Region based hidden markov random field model for brain mr image segmentation. *World Academy of Science Eng. Technol.*, 4:233–236, 2005.

[10] J. Cohen-Adad, H. Benali, R.D. Hoge, and S. Rossignol. In vivo DTI of the healthy and injured cat spinal cord at high spatial and angular resolution. *NeuroImage*, 40(2):685–697, Apr 2008.

[11] Cynthia A. DeBoy, Jiangyang Zhang, Sonny Dike, Irina Shats, Melina Jones, Daniel S. Reich, Susumu Mori, Thien Nguyen, Brian Rothstein, Robert H. Miller, John T. Griffin, Douglas A. Kerr, and Peter A. Calabresi. High resolution diffusion tensor imaging of axonal damage in focal inflammatory and demyelinating lesions in rat spinal cord. *Brain*, 130(8):2199–2210, 2007.

[12] A. Degrassi, M. Russo, C. Nanni, V. Patton, R. Alzani, A. M. Giusti, S. Fanti, M. Ciomei, E. Pesenti, and G. Texido. Efficacy of PHA-848125, a cyclin-dependent kinase inhibitor, on the K-Ras(G12D)LA2 lung adenocarcinoma transgenic mouse model: evaluation by multimodality imaging. *Mol. Cancer Ther.*, 9(3):673–681, 2010.

[13] Arthur Dempster, Nan Laird, and Donald Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38, 1977.

[14] Xiang Deng, Jaivijay Ramu, and Ponnada A. Narayana. Spinal cord atrophy in injured rodents: High-resolution MRI. *Magn. Reson. Med.*, 57(3):620–624, 2007.

[15] Therese A. Dolecek, Jennifer M. Propp, Nancy E. Stroup, and Carol Kruchko. CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 20052009. *Neuro Oncol*, 4 (suppl 5):v1–v49, 2012.

[16] Benjamin M. Ellingson, John L. Ulmer, and Brian D. Schmit. Gray and white matter delineation in the human spinal cord using diffusion tensor imaging and fuzzy logic. *Acad. Radiol.*, 14(7):847–858, July 2007.

[17] Benjamin M. Ellingson, John L. Ulmer, and Brian D. Schmit. Morphology and morphometry of human chronic spinal cord injury using diffusion tensor imaging and fuzzy logic. *Ann. Biomed. Eng.*, 36(2):224–236, Feb 2008.

[18] Jeffrey A. Engelman, Liang Chen, Xiaohong Tan, Katherine Crosby, Alexander R. Guimaraes, Rabi Upadhyay, Michel Maira, Kate McNamara, Samanthi A. Perera, Youngchul Song, Lucian R. Chirieac, Ramneet Kaur, Angela Lightbown, Jessica Simendinger, Timothy Li, Robert F. Padera, Carlos García-Echeverría, Ralph Weissleder, Umar Mahmood, Lewis C. Cantley, and Kwok-Kin Wong. Effective use of PI3K and MEK inhibitors to treat mutant Kras G12D and PIK3CA H1047R murine lung cancers. *Nat. Med.*, 14(12):1351–1356, 2008.

[19] Zvi G. Fridlender, Veena Kapoor, George Buchlis, Guanjun Cheng, Jing Sun, Liang-Chuan S. Wang, Sunil Singhal, Linda A. Snyder, and Steven M. Albelda. Monocyte chemoattractant protein-1 blockade inhibits lung cancer tumor growth by altering macrophage phenotype and activating CD8+ cells. *Am. J. Respir. Cell Mo. Biol.*, 44:230–237, 2010.

[20] Joel R. Garbow and Joseph J. H. Ackerman. Imaging primary lung cancers in mice to study radiation biology: In regard to kirsch et al. (int j radiat oncol biol phys 2010;76:973-977). *Int. J. Radiat. Oncol. Biol. Phys.*, 79:959, 2010.

[21] Joel R. Garbow, Joseph P. Dugas, Sheng-Kwei Song, and Mark S. Conradi. A simple, robust hardware device for passive or active respiratory gating in MRI and MRS experiments. *Concepts Magn. Reson. B Magn. Reson. Eng.*, 21B:40–48, 2004.

[22] Joel R. Garbow, Min Wang, Yian Wang, Ronald A. Lubet, and Ming You. Quantitative monitoring of adenocarcinoma development in rodents by magnetic resonance imaging. *Clin. Cancer Res.*, 14(5):1363–1367, 2008.

[23] Joel R. Garbow, Zhongqiu Zhang, and Ming You. Detection of primary lung tumors in rodents by magnetic resonance imaging. *Cancer Res.*, 64(8):2740–2742, 2004.

[24] Carlos F. G. C. Geraldes and Sophie Laurent. Classification and basic properties of contrast agents for magnetic resonance imaging. *Contrast Media Mol. Imaging*, 4(1):1–23, 2009.

[25] Robert T. Greenlee, Taylor Murray, Sherry Bolden, and Phyllis A. Wingo. Cancer statistics. *CA Cancer J. Clin.*, 50:7–33, 2000.

[26] Jayaroop Gullapalli, Jaroslaw Krejza, and Eric D. Schwartz. *In vivo* DTI evaluation of white matter tracts in rat spinal cord. *J. Magn. Reson. Imaging*, 24(1):231–234, 2006.

[27] L. B. Jakeman, Z. Guan, P. Wei, R. Ponnappan, R. Dzwonczyk, P. G. Popovich, and BT Stokes. Traumatic spinal cord injury produced by controlled contusion in mouse. *J. Neurotrauma*, 17(4):299–319, Apr 2000.

[28] Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, and Stephen M. Smith. FSL. *NeuroImage*, 62(2):782–790, 2012.

[29] Derek R. Johnson and Brian Patrick ONeill. Glioblastoma survival in the united states before and during the temozolomide era. *J Neurooncol*, 107:359–364, 2012.

[30] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *Int'l J. of Computer Vision*, 1(4):259–268, 1987.

[31] Artem Khmelinskii, Martin Baiker, Eric L. Kaijzel, Josette Chen, Johan H. C. Reiber, and Boudewijn P. F. Lelieveldt. Articulated whole-body atlases for small animal image analysis: Construction and applications. *Mol. Imaging Biol.*, 13(5):898–910, 2011.

[32] Joong Hee Kim, D.N. Loy, Hsiao-Fang Liang, K. Trinkaus, R.E. Schmidt, and Sheng-Kwei Song. Noninvasive diffusion tensor imaging of evolving white matter pathology in a mouse model of acute spinal cord injury. *Magn. Reson. Med.*, 58(2):253–260, Aug 2007.

[33] David G. Kirsch, Jan Grimm, Alexander R. Guimaraes, Gregory R. Wojtkiewicz, Bradford A. Perez, Philip M. Santiago, Nikolas K. Anthony, Thomas Forbes, Karen Doppke, Ralph Weissleder, and Tyler Jacks. Imaging primary lung cancers in mice to study radiation biology. *Int. J. Radiat. Oncol. Biol. Phys.*, 76:973–977, 2010.

[34] Cheng Guan Koay, Lin-Ching Chang, John D. Carew, Carlo Pierpaoli, and Peter J. Basser. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *J. Magn. Reson.*, 182:115–125, 2006.

[35] Natasa Kovacevic, Ghassan Hamarneh, and Mark Henkelman. Anatomically guided registration of whole body mouse MR images. *Lect. Notes Comput. Sc.*, 2879:870–877, 2003.

[36] B. P. F. Lelieveldt, R. J. van der Geest, M. Ramze Rezaee, J. G. Bosch, and J. H. C. Reiber. Anatomical model matching with fuzzy implicit surfaces for segmentation of thoracic volume scans. *IEEE Trans. Med. Imaging*, 18(3):218–230, 1999.

[37] Boudewijn P. F. Lelieveldt, Milan Sonka, Lizann Bolinger, Thomas D. Scholz, Hein Kayser, Rob van der Geest, and Johan H. C. Reiber. Anatomical modeling with fuzzy implicit surface templates: Application to automated localization of the heart and lungs in thoracic MR volumes. *Comput. Vis. Image Und.*, 80(1):1–20, 2000.

[38] P. C. A. Louis. Medical statistics. *Am. J. Med. Sci.*, 21:525–528, 1837.

[39] D.N. Loy, Joong Hee Kim, M. Xie, R.E. Schmidt, K. Trinkaus, and Sheng-Kwei Song. Diffusion tensor imaging predicts hyperacute spinal cord injury severity. *J. Neurotrauma*, 24(6):979–990, Jun 2007.

[40] M. Ma, D. M. Basso, P. Walters, B. T. Stokes, and L. B. Jakeman. Behavioral and histological outcomes following graded spinal cord contusion injury in the C57BL/6 mouse. *Exp. Neurol.*, 169(2):239–254, Jun 2001.

[41] D.S. Magnuson, R. Lovett, C. Coffee, R. Gray, Y. Han, Y.P. Zhang, and D.A. Burke. Functional consequences of lumbar spinal cord contusion injuries in the adult rat. *J. Neurotrauma*, 22(5):529–543, May 2005.

[42] L. Martiniova, M. S. Kotys, D. Thomasson, D. Schimel, E. W. Lai, M. Bernardo, M. J. Merino, J. F. Powers, J. Ruzicka, R. Kvetnansky, P. L. Choyke, and K. Pacak. Noninvasive monitoring of a murine model of metastatic pheochromocytoma: a comparison of contrast-enhanced microCT and nonenhanced MRI. *J. Magn. Reson. Imaging*, 29(3):685–691, 2009.

[43] M.L. McEwen and J.E. Springer. Quantification of locomotor recovery following spinal cord contusion in adult rats. *J. Neurotrauma*, 23(11):1632–1653, Nov 2006.

[44] Ian Middleton and Robert I. Damper. Segmentation of magnetic resonance images using a combination of neural networks and active contour models. *Med. Eng. Phys.*, 26:71–86, 2004.

[45] J. D. Minna, J. A. Roth, and A. F. Gazdar. Focus on lung cancer. *Cancer Cell*, 1:49–52, 2002.

[46] J. A. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, 7:308–313, 1965.

[47] Stanley Osher and James A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. of Comput. Phys.*, 79:12–49, 1988.

[48] Paul M. Parizel, Luc van den Hauwe, Frank De Belder, J. Van Goethem, Caroline Venstermans, Rodrigo Salgado, Maurits Voormolen, and Wim Van Hecke. Magnetic resonance imaging of the brain. In P. Reimer, P. M. Parizel, J. F. M. Meaney, and F.-A. Stichnoth, editors, *Clinical MR Imaging*, pages 107–195. Springer Berlin Heidelberg, New York, 2010.

[49] Nilanjan Ray, Scott T. Acton, Talissa Altes, Eduard E. de Lange, and James R. Brookeman. Merging parametric active contours within homogeneous image regions for MRI-based lung segmentation. *IEEE Trans. Med. Imaging*, 22(2):189–199, 2003.

[50] Y. Rosen, G. Ramniceanu, R. Margalit, D. Grobgeld, R. Eilam, H. Degani, and E. Furman-Haran. Vascular perfusion of human lung cancer in a rat orthotopic model using dynamic contrast-enhanced magnetic resonance imaging. *Int. J. Cancer*, 119(2):365–372, 2006.

[51] Stephen W. Scheff, Alexander G. Rabchevsky, Isabella Fugaccia, John A. Main, and James E. Lumpp. Experimental modeling of spinal cord injury: Characterization of a force-defined injury device. *J. Neurotrauma*, 20(2):179–193, Feb 2003.

[52] E.D. Schwartz, J. Duda, J.S. Shumsky, E.T. Cooper, and J. Gee. Spinal cord diffusion tensor imaging and fiber tracking can identify white matter tract disruption and glial

scar orientation following lateral funiculotomy. *J. Neurotrauma*, 22(12):1388–1398, Dec 2005.

[53] William F. Sensakovic, Samuel G. Armato III, Adam Starkey, and Philip Caligiuri. Automated lung segmentation of diseased and artifact-corrupted magnetic resonance sections. *Med. Phys.*, 33(9):3085–3093, 2006.

[54] Natalie J. Serkova, Zachary Van Rheen, Meghan Tobias, Joshua E. Pitzer, J. Erby Wilkinson, and Kathleen A. Stringer. Utility of magnetic resonance imaging and nuclear magnetic resonance-based metabolomics for quantification of inflammatory lung injury. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, 295:L152–L161, 2008.

[55] Bronwyn Stall, Leor Zach, Holly Ning, John Ondos, Barbara Arora, Uma Shankavaram, Robert W. Miller, Deborah Citrin, and Kevin Camphausen. Comparison of t2 and flair imaging for target delineation in high grade gliomas. *Radiat. Oncol.*, 5:5, 2010.

[56] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.*, 42(1):288–292, Jan 1965.

[57] Janakiraman Subramanian, Dainel Morgensztern, Boone Goodgame, Maria Q. Baggstrom, Feng Gao, Jay Piccirillo, and Ramasqamy Govindan. Distinctive characteristics of non-small cell lung cancer (NSCLC) in the young: A surveillance, epidemiology, and end results (SEER) analysis. *J. Thorac. Oncol.*, 5(1):23–28, 2010.

[58] C.-H. Toh, M. Castillo, A.M.-C. Wong, K.-C. Wei, H.-F. Wong, S.-H. Ng, and Y.-L. Wan. Primary cerebral lymphoma and glioblastoma multiforme: Differences in diffusion characteristics evaluated with diffusion tensor imaging. *Am. J. Neuroradiol.*, 29:471–475, 2009.

[59] U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 1999-2006 Incidence and Mortality Web-based Report.* Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, 2010.

[60] W. M. Wells, E. L. Grimson, R. Kikinis, and F. A. Jolesz. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imaging*, 15(4):429–442, 1996.

[61] A. A. Younis, N. Ramirez, P. M. Pattany, R. J. Burns, and M. I. Sharawy. Automated segmentation of spinal diffusion tensor MR imaging. *SoutheastCon, 2005. IEEE Proc.*, pages 187–192, April 2005.

[62] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging*, 20(1):45–57, 2001.

[63] X. Zhou, H. Bao, R. Al-Hashem, H. Ji, M. Albert, K. K. Wong, and Y. Sun. Magnetic resonance imaging of the response of a mouse model of non-small cell lung cancer to tyrosine kinase inhibitor treatment. *Comp. Med.*, 58(3):276–281, 2008.

Publications List

[1] Alexander Sasha Krupnick / Vanessa K. Tidwell, John A. Engelbach, Vamsi V. Alli, Arye Nehorai, Ming You, Haris G. Vikis, Andrew E. Gelman, Daniel Kreisel, and Joel Garbow. Quantitative monitoring of mouse lung tumors by magnetic resonance imaging. *Nat. Protoc.*, 7(1):128–142, 2012.

[2] Vanessa K. Tidwell, Joel Garbow, Alexander Sasha Krupnick, John A. Engelbach, and Arye Nehorai. Quantitative analysis of tumor burden in mouse lung via mri. *Magn. Reson. Med.*, 67(2):572–579, 2012.

[3] Vanessa K. Tidwell, Joong H. Kim, Sheng-Kwei Song, and Arye Nehorai. Automatic segmentation of rodent spinal cord diffusion mr images. *Magn. Reson. Med.*, 64(3):893–901, 2010.