A New Image Quantitative Method for Diagnosis and Therapeutic Response

by

Wenzhe Xue

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2016 by the
Graduate Supervisory Committee:

David Kaufman, Co-Chair
J. Ross Mitchell, Co-Chair
Matthew Scotch
William Johnson

ARIZONA STATE UNIVERSITY

December 2016

ABSTRACT

Accurate quantitative information of tumor/lesion volume plays a critical role in diagnosis and treatment assessment. The current clinical practice emphasizes on efficiency, but sacrifices accuracy (bias and precision). In the other hand, many computational algorithms focus on improving the accuracy, but are often time consuming and cumbersome to use. Not to mention that most of them lack validation studies on real clinical data. All of these hinder the translation of these advanced methods from benchside to bedside.

In this dissertation, I present a user interactive image application to rapidly extract accurate quantitative information of abnormalities (tumor/lesion) from multi-spectral medical images, such as measuring brain tumor volume from MRI. This is enabled by a GPU level set method, an intelligent algorithm to learn image features from user inputs, and a simple and intuitive graphical user interface with 2D/3D visualization. In addition, a comprehensive workflow is presented to validate image quantitative methods for clinical studies.

This application has been evaluated and validated in multiple cases, including quantifying healthy brain white matter volume from MRI and brain lesion volume from CT or MRI. The evaluation studies show that this application has been able to achieve comparable results to the state-of-the-art computer algorithms. More importantly, the retrospective validation study on measuring intracerebral hemorrhage volume from CT scans demonstrates that not only the measurement attributes are superior to the current practice method in terms of bias and precision but also it is achieved without a significant delay in acquisition time. In other words, it could be useful to the clinical trials and clinical practice, especially when intervention and prognostication rely upon accurate baseline lesion volume or upon detecting change in serial lesion

volumetric measurements. Obviously, this application is useful to biomedical research areas which desire an accurate quantitative information of anatomies from medical images. In addition, the morphological information is retained also. This is useful to researches which require an accurate delineation of anatomic structures, such as surgery simulation and planning.

To my beloved wife, Xuelian

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my advisors. Dr. Mitchell, thank you for the endless support on my research. Your passion to the science and technology highly influenced me. Dr. Kaufman, thank you for your help and guidance on my Ph.D study. I would like to thank the rest of my dissertation committee: Dr. Scotch and Dr. Johnson. You widened my research from various perspectives.

Thank you to the doctors in the Mayo Clinic who contributed to my research in many way: Dr. Zwart, Dr. Silva, Dr. Hoxworth, Dr. Wingerchuk, Dr. Aguilar, Dr. Patel, Dr. Demaerschalk, for providing insight, labeling data, and coauthoring the papers. Also, thank you Dr. Swanson for your generosity.

I would like to thank the staffs in the department of Biomedical Informatics. Thanks in particular to Maria and Lauren who helped to coordinate my academic study plan.

I thank my lab mates in both Medical Image Informatics lab and Precision Neurotherapeutic lab: Sara, Simon, Cyprien, Pyari, Gil, Helene, Kun, Stefanie, Fillan, Pamela, Susan, and Andrew.

Last but not the least, I would like to thank my family. Thank you to my selfless and supportive parents. Also thank you Xinnong for inspiring me to study in US. And most of all, thank you to my wife Xuelian. You have always been with me during this long journey. I look forward to the many years to come.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Healthcare generates tremendous amounts of data everyday. These data range from genetics to tissues and organs, from one person to a large population. How to effectively use this data to extract information and obtain knowledge for better analysis and decision making becomes the main question of biomedical informatics. Among all of the data, a large portion is generated from modern medical imaging modalities. Medical images enable physicians to peer through the patient's body non-invasively. Various medical imaging techniques cover every part of the human body to provide both structural and functional information. Biomedical imaging informatics is an active research field which deals with medical imaging data. The field of biomedical imaging informatics encompasses multiple areas in the medical imaging workflow including digital image acquisition, image content representation, image management/storage, image processing, and image interpretation and computer reasoning.

My research belongs to the image processing step, which extracts information from medical images to assist decision making and image analysis in the next step. Currently, qualitative information is mostly used in the clinical practice. What really wanted is quantitative information, which are numerical values that can be compared and computed. One simple and common quantitative information from medical images is size. Here are several examples of how size could be used in either clinical practice and biomedical research: sizes of the brain over time show brain atrophy, which is associated with either normal aging or neurological disease, such as Alzheimer's

disease; sizes of lesions such as multiple sclerosis lesion is used to tell the stage of disease; change of hemorrhage sizes within hours indicates the speed of hematoma growth, which is a strong predictor of patient care outcome.

However, most of the information is qualitative in clinical practice. A survey on the radiology charts from a hospital showed that less than half of the charts contain any quantitative metrics while only less than 3% use advanced quantitative metrics. The measurement of size is even less. In other words, this is rarely used in practice. In clinical trials, diameter based methods are normally used which estimates the volume of an object using one or two axes in millimeters. It is simple and rapid to use, but suffers high bias and variability. In order to achieve acceptable results, manual outlining by experts such as radiologists or trained technicians is often used. But manual outlining is very time consuming and not clinically feasible. It is conducted mostly in biomedical research which requires to track object size over time. On the other hand, there are image segmentation algorithms which perform similar tasks. My research uses the level set method, which requires the user to place seed points to grow and evolves iteratively to extract the targeted object. It has been proven to obtain acceptable results from medical images but is neither simple nor rapid. Previous effort had accelerated level set method with graphical processing units (GPU), which solved part of the problem why level set method is slow. However several parameters which need to be specified by the user hinder the simplicity and rapidness of level set method. The long term goal is to develop a quantitative image method to measure size from medical images that satisfy three criteria: 1) simple, 2) rapid, and 3) accurate (low bias and variability).

In this dissertation, I present this amazing tool we developed to extract quantitative information from medical images and its evaluation and validation on brain white

matter, multiple sclerosis lesion, and intracerebral hemorrhage stroke. These studies show that this toolkit has minimal user interaction and intuitive graphical user interface (simple), is on par with the less accurate measurement methods used in clinical practice or trials (rapid), and has small bias and low variability (accurate).

This application provides a simple, rapid, and reliable way to extract quantitative information from medical images. This information is very useful to assist further analysis and better decision making. It could be used to extract quantitative information from CT or MRI data which provides a high resolution of structural informations. For instance, it could be used to track the size change of solid tumors, which is related to cancer research and treatment response. In addition, it could be used to measure organ volume, like liver or kidney, to aid transplant surgeries. But there are some limitations. This application is not reliable when using low resolution functional images, such as functional MRI data. Another limitation is brought by the low contrast between targeted object and surrounding regions.

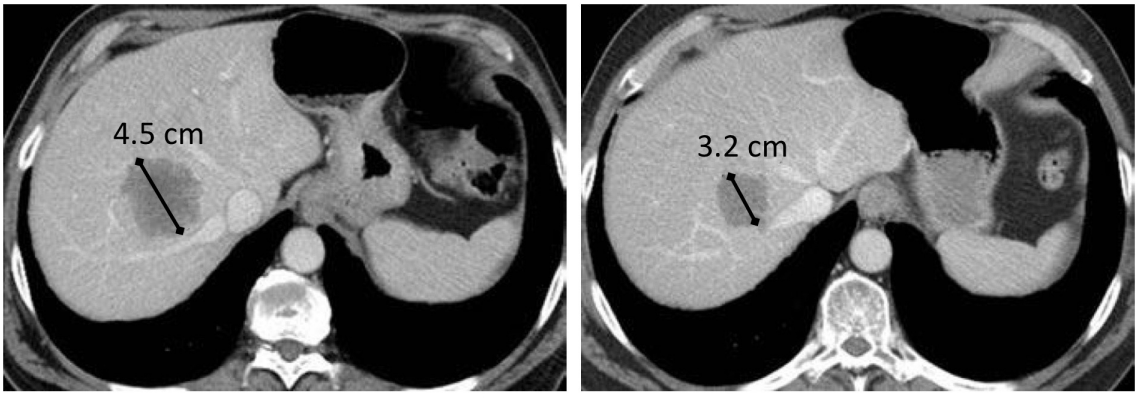

Figure 1. A demonstration of using diameter based method to measure liver tumors (the dark region) inside the liver (light gray area). The left figure shows the tumor size (4.5 cm) from the baseline CT scan before treatment . The right figure shows the tumor size (3.2 cm) from the follow up CT scan after treatment.

Chapter 2 covers the background of this research, from biomedical imaging informatics to image segmentation algorithms using commodity graphics card. Then, Chapter 3 and 4 present the developments of our method, applying to extract brain white matter region and multiple sclerosis lesions (brain lesions) from MR images. Chapter 5 presents a comprehensive clinical validation study to evaluate the bias, precision, and measurement time for our method to quantify intracerebral hemorrhage volume from CT images. Chapter 6 discusses the impact of our method to both clinical practice and researches. In the end, Chapter 7 concludes this dissertation.

Chapter 2

BACKGROUND

2.1 Biomedical Imaging Informatics

Biomedical imaging informatics is a subdiscipline of biomedical informatics. As a fast evolving research field at the crossing of medicine, informatics, and computer science and engineering, it plays a critical role in clinical practices, education, and research. The field of biomedical imaging informatics encompasses multiple areas in the medical image workflow including digital image acquisition, image content representation, image management/storage, image processing, image interpretation and computer reasoning (Shortliffe and Cimino, 2006). For instance, a patient with severe pain in the arm is prescribed an X-ray of his/her arm (digital image acquisition). The X-ray image is stored as a digital format into the radiology image database and management system (image content representation and image management/storage). This image could be enhanced by adjusting brightness and contrast (image processing) for better and easier examination. After the image is ready to review, the radiologist opens the image from his workstation, identifies a bone fracture, potentially aided by computer software (image interpretation/computer reasoning).

There are two strategies in imaging the body. The first is structure imaging, delineating anatomic structure. Structure imaging captures a high spatial resolution image to accurately represent the structure of the body, including healthy organs and abnormalities. Both CT and MRI are popular structure imaging modalities with high spatial resolution. Computed Tomography (CT) is cross sectional and volumetric

images, reconstructed from the radiation attenuation of a series of X-ray projected from various angels throughout the human body (Buzug, 2008). The intensity at each voxel (volume element), representing the density of the tissue in that position, is estimated using computer reconstruction algorithms. Magnetic Resonance Imaging (MRI) uses the body's natural magnetic properties to produce detailed images from any part of the body (Berger, 2002).

The second strategy is functional imaging, determines tissue composition or function, which has been an active research area recent years, especially functional brain imaging modalities. Positron Emission Tomography (PET) shows high functional information but low spatial resolution (Mettler Jr and Guiberteau, 2011). It uses positron emitters for producing the image. In addition, new technologies have been in development to combine structure and functional imaging. For instance, PET/CT combines the characteristics of PET and CT images to visualize low resolution functional PET scan on top of high spatial resolution CT scan. Presently, most clinical studies use a combination of modalities to generate both structural and functional information.

Medical image processing The increasingly prominent role of medical imaging in medical diagnosis and treatment leads to the challenging problem of extracting clinically useful information about anatomic structures from modern medical images using computers. Although modern imaging devices provide high-resolution views of internal anatomy, the use of computers to quantitatively analyze anatomic structures with reliability and efficiency is limited. Accurate and quantitative information must be efficiently extracted in order to support clinical practice and research. Like all digital data, medical images can be manipulated and processed with computational algorithms. There are a broad variety of image processing methods including: transformations

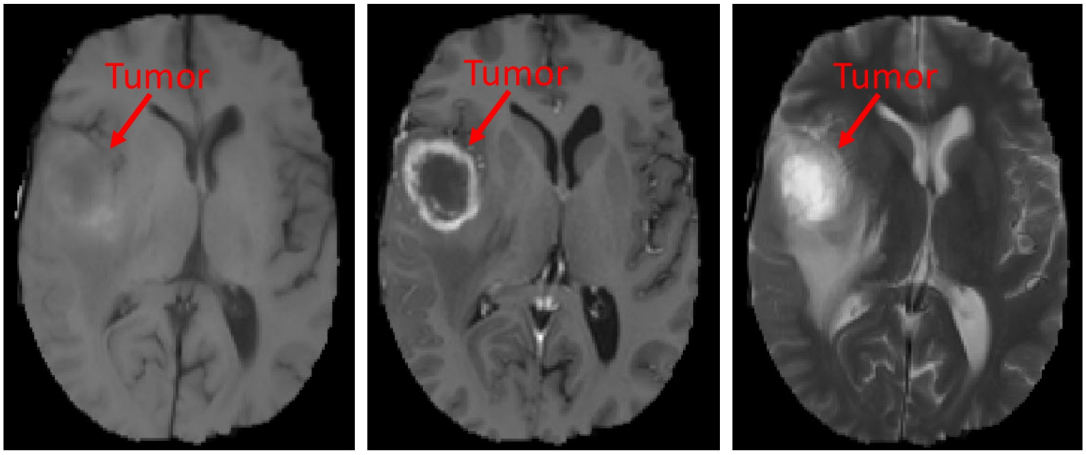

Figure 2. Images of the brain from three modalities (Left: T1 without contrast, middle: T1 with contrast, and right: T2). A brain tumor is located in the left side of the image.

to enhance visualization, computations to extract features, and systems to automate detection or diagnose abnormalities in the images. When applying these image processing methods to medical images, it faces unique and different challenges even though it uses similar methods of general image processing. One of the differences is that these data are 3D volumetric data. Since most of the images are from multi-modalities (CT, MRI, PET), they are treated as a multi-spectral images with more channels than regular images with only three channels (Red, Green, Blue). In addition, each modality has its own unique physical characteristics, thus causing unique noise, contrast and other issues. The fusion/co-registration of information across several modalities is a challenge that needs to be addressed as well.

2.2 Medical Image Segmentation

Image segmentation is one of the most challenging and researched tasks in image processing. It is employed as one of the most fundamental processes in extracting

quantitative and geometry information of a region of interest (ROI) from medical image data, and it determines the quality of the healthcare outcome (Fitzpatrick and Sonka, 2000; Demirkaya *et al.*, 2009). The goal of medical image segmentation is to automatically or semi-automatically delineate anatomic structures and other ROI from medical images. Then quantitative information and characteristics of the ROIs can be extracted.

One common scenario to obtain accurate tumor volume is manually delineating the tumor boundary in 2D axial slices by trained experts, which are later reconstructed to 3D surface by software, such as OsiriX (Rosset *et al.*, 2004). The manual method is often used to create ground truth, precise and accurate but time consuming, for validation of computer algorithms. Researchers have proposed many automatic and semi-automatic methods for the segmentation of anatomical structures from medical images using computers.

- Gray-level thresholding is the early and simple segmentation method. In this approach, we assign a label to objects by comparing their gray-level value to one or more intensity thresholds (Fitzpatrick and Sonka, 2000). It is often effective for images in which different structures have high contrast intensities but difficult to handle image noise.
- Parametric Deformable models (McInerney and Terzopoulos, 1996), such as Snake (Kass *et al.*, 1988; Xu and Prince, 1998) are model-based techniques for delineating region boundaries using closed parametric curves or surfaces under the influence of internal and external forces. The level set method is a non-parametric deformable model (Osher and Sethian, 1988a; Malladi *et al.*, 1995; Caselles *et al.*, 1993). Deformable models provide the user with efficient interactivity. But, in order to detect the boundary, a curve or surface must be

placed near the desired boundary before the iterative process. In addition, it requires longer computational time to achieve acceptable results.

- Active shape models statistically learn the variance of the object's shape, and constrain its topology and general shape to obtain the segmentation (Cootes *et al.*, 1995). In addition, active appearance models (Cootes *et al.*, 1998) incorporate appearance priors (intensity value) and shape. These methods are useful to extract anatomic structures with similar shapes, such as the liver and lung. But they may not be effective to segment objects with large shapes variation.
- Live wire (Falcao *et al.*, 1998; Mortensen and Barrett, 1998) and fuzzy connectedness methods (Udupa and Samarasekera, 1996; Saha and Udupa, 2000) are semi-automated segmentation methods that are based on finding optimal paths with dynamic programming techniques. These methods treat the images as graphs. These methods are easy and intuitive for the use and have the added advantage of providing the results in real time. But they are also sensitive to noise; the user cannot adjust the smoothness of the boundary.
- Image segmentation with graph cut (Boykov and Jolly, 2001; Boykov and Kolmogorov, 2004) partitions an image into foreground and background by computing a globally optimal cut. It requires users to select seed points for both foreground and background. Grabcut (Rother *et al.*, 2004) extends graph cut to an iterative process. Random walk (Grady, 2006) is another graph-based image segmentation algorithm. These methods are computationally expensive for high dimensional images.

A comparison of computer assisted image segmentation algorithms is shown in Table 2.2.

Table 1. Comparison of Advanced Image Segmentation Methods

Approach	Pros	Cons
Deformable model	1) Easily manipulated using external image forces. 2) Can be used to track dynamic objects in temporal and spatial dimensions. 3) Users are able to guide the snake with mouse actions.	1) Have to place contour near the desired boundary. 2) Sensitive to noises and local minimum. 3) Cannot merge or split contour. 4) Longer computational time in order to achieve more accurate results.
Livewire	1) Easy and intuitive for the use 2) Show the result in real-time. 3) Easy to implement.	1) Sensitive to noise. 2) Cannot adjust smoothness. 3) Require more user interaction for more difficult tasks.
Graph Cut	1) Fast inference. 2) Incorporate recognition or high-level priors. 3) Easily extend to 3D image segmentation.	1) Only apply to binary segmentation. 2) No uncertainty measure associated with the solution. 3) Leakage on weak edges.
Random Walk	1) Segment multiple objects. 2) Robust on weak edges. 3) Return probabilistic value to the final segmentation 4) Only one parameter	Computationally expensive for high dimensional images.
Level Set	1) Implicit surface representation. 2) Manipulate less parameters of the surface or contour. 3) Easily follow the topological shape changes, such as split, merge, *ect.*	1) Time consuming since solving a partial differential equation every iteration. 2) Difficult to prevent leakage through weak boundaries.

2.2.1 Level Set Method

The level set method is popular for 3D medical image segmentation due to its robustness on 1) handling noisy images and 2) handling topological changes such as merging and splitting. In this section, I discuss this method in detail because it serves as a foundation for my research.

The level set method embeds an implicit surface within an image, and iteratively deforms the surface to envelop the containing ROI. To better explain the concept, here is an instance of a 2D case (Figure 3). In an 2D image, the level set method represents the 2D contour with the zero level set of a 3D surface. The propagation of the contour becomes the 3D surface evolution. Each point on this 3D surface moves along the direction which is perpendicular to the local surface. And how fast it moves is determined by a speed function based on the image features and local surface curvature. The algorithms begin with a user-placed seed region of interest (ROI). Then the ROI is deformed iteratively to encompass the entire ROI, in our scenario, the tumor or lesion. In addition to the seed region, several adjustment parameters (related to image intensities and surface smoothness) control the growth and smoothness of the ROI surface and can be specified by the user.

The implicitly represented level set surface is defined as $\{\mathbf{x}|\phi(\mathbf{x},t)=0\}$, where $\mathbf{x}$ is a coordinate in the image volume, t is the current iteration time in the level set evolution, and $\phi(\mathbf{x},t):\Re^4\mapsto\Re$ refines the level set according to:

$$\phi(\mathbf{x},t)=\phi(\mathbf{x},t-\Delta t)+\nabla t\cdot F(\mathbf{x},t)|\nabla\phi(\mathbf{x},t-\Delta t)|. \tag{2.1}$$

A speed function $F(\mathbf{x})$ defines the rate of motion of each local point on the implicit surface. The deforming direction of each point is along the norm of the local surface.

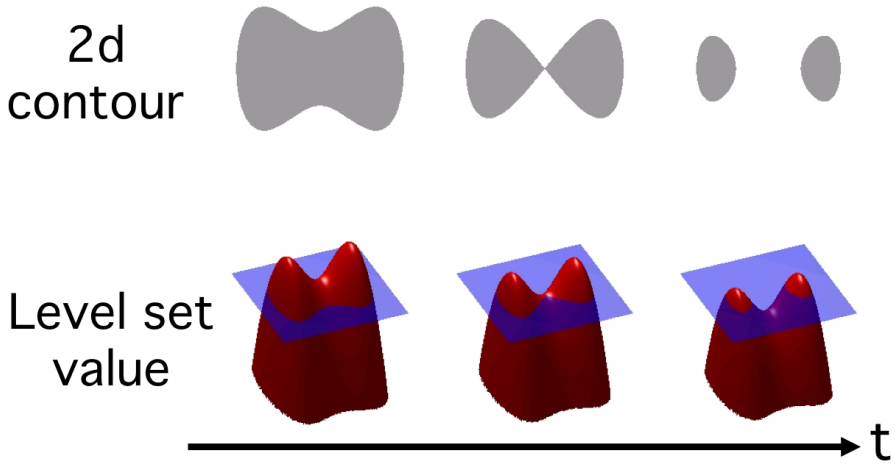

Figure 3. An illustration of the level set method in a 2D image. The top row shows the 2D shape change and split. The bottom row shows the corresponding change of the implicit 3D level set surface (red surface). The blue plane is the zero level set.

We adopted the speed function proposed by Lefohn *et al.*:

$$F(\mathbf{x}, t) = \alpha C(\mathbf{x}, t) + (1 - \alpha) D(\mathbf{x}), \tag{2.2}$$

where $C(\mathbf{x}, t)$ is the curvature term, $D(\mathbf{x})$ is the data term, and $\alpha \in [0, 1]$, the weight of curvature term, is a blending term that controls the relative contribution of the curvature and data terms.

The data term used in Lefohn *et al.* is a function of intensities in a single image volume:

$$D(\mathbf{x}) = \epsilon - (|I(\mathbf{x})| - T), \tag{2.3}$$

where $I(\mathbf{x})$ is the image intensity at location $\mathbf{x}$, T is a user-specified target intensity that encourages maximum level set growth, and ϵ is a user-specified parameter that indicates the range of intensities around T that will promote level set growth. If $T - \epsilon < I(\mathbf{x}) < T + \epsilon$, then $D(\mathbf{x})$ will promote surface growth. Otherwise it will promote surface contraction.

The curvature term $C(\mathbf{x}, t)$ depends upon the mean curvature of the local surface from the previous iteration:

$$C(\mathbf{x}, t) = \nabla \cdot \frac{\nabla\phi(\mathbf{x}, t - \Delta t)}{|\nabla\phi(\mathbf{x}, t - \Delta t)|}. \tag{2.4}$$

A degree of curvature prevents critical leakage through weak boundaries and fills holes caused by noise inside the segmented region. Excessive smoothing by overly penalizing curvature can significantly distort segmented objects. Eventually, when the weight of the curvature term α is too large (close to 1), the level set surface will shrink to a point.

This approach presents unique advantage over the parametric deformable model. The surface represented by the level set function handles the topological changes naturally by seamlessly developing sharp corners, breaking or merging during the evolution. It has been proven that level set method is very useful with three or higher dimensional segmentation problems as the surfaces become inherently complex for parametric representation (Sethian, 1999). However, level set segmentation has been too computationally expensive and time consuming for widespread clinical practice.

Graphics processing units (GPU) was originally designed to process and render computer graphics. GPU's architecture is composed of thousands of small processors in order to process computational tasks for pixels, while central processing unit (CPU) only uses several powerful processing cores. Thanks to the boom of video game industry, GPUs have been getting better and better with more computational powers, larger memory bandwidth, and increasing programmability. In the past ten years, GPUs become extremely popular for scientific computing which are often computational expensive. They significantly accelerate applications in the fields ranging from artificial intelligence to robotics, from computer vision to big data analysis.

2.3 General Purpose Computation Using Graphics Processing Units (GPGPU)

In the last decade, researchers have started to deploy GPU as a massively parallel platform to accelerate general computations. General purpose computation on graphic processing units (GPGPU) employs commodity graphical hardware, which typically render computer graphics, to perform computations traditionally handled by the central processing unit (CPU) (Eklund *et al.*, 2013). With increasing programmability and higher precision arithmetic processing (Figure 4), GPUs demonstrate potential as cost-effective and high performance computing platforms for various applications. Medical image processing also benefit from the parallelism of GPUs. Medical imaging applications (NVIDIA, 2012), such as image registration, high-dimensional image denoising, and image segmentation (Figure 5). It could enable practical use of computationally demanding algorithms, such as level set method.

2.4 GPU Level Set Segmentation

Lefohn *et al.* started a GPU-based interactive level set segmentation in 2003 and showed the potential of GPU acceleration on narrowband level set method, which solves the problem by limiting the computation in a thin band around the propagating surface (Lefohn *et al.*, 2003a,b, 2004). Cates *et al.* integrated Lefohn's algorithm and implemented a tool called GIST for 3D medical image segmentation (Cates *et al.*, 2004). More recently, Roberts *et al.* developed a new GPU level set algorithm, which is 16x faster than the previously fastest GPU level set (Roberts *et al.*, 2010). First, this method limits the active computational domain to the minimal set of changing

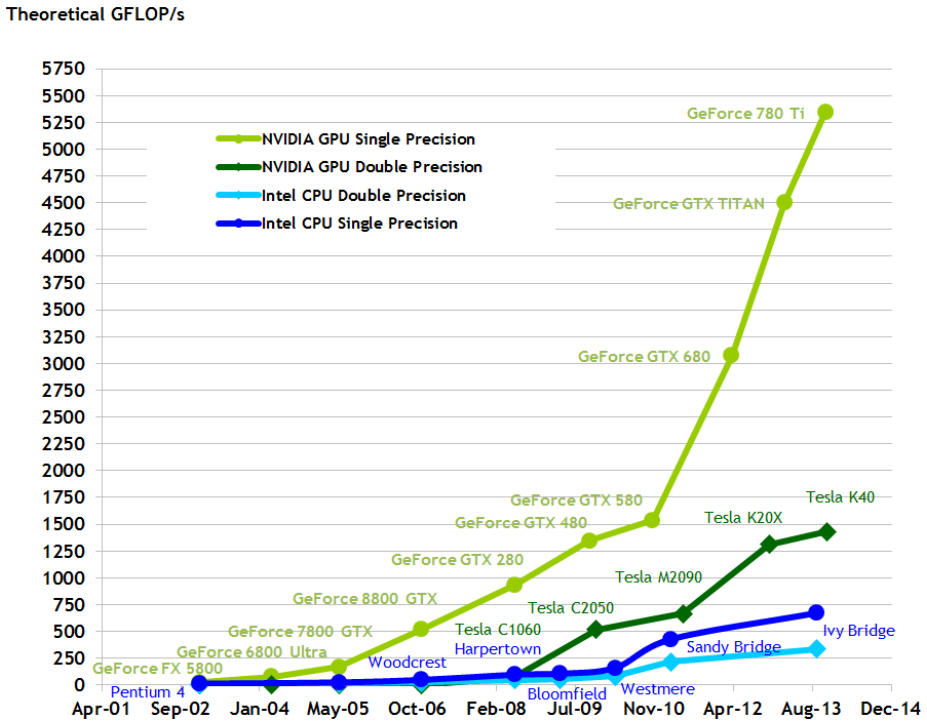

Figure 4. Floating-Point Operations per Second for the CPU and GPU. (CUDA programming guide, 2013)

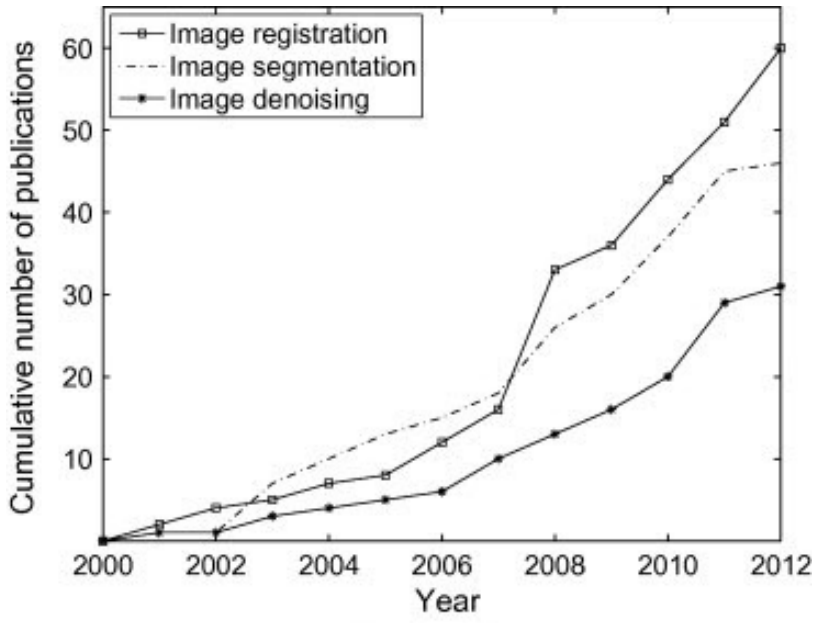

Figure 5. The figure shows the cumulative number of publications for GPU-based medical imaging algorithms (Eklund *et al.*, 2013).

elements by examining both the temporal and spatial derivative of the level set field. Second, it is a work- and step-efficient *implementation of this algorithm for GPU architectures. These features minimize the size of the active computational domain to remain small as the algorithm progresses. Meanwhile, it requires less memory than previous GPU algorithms. These innovations allow the algorithm to work quickly on very large image volumes. In addition, GPU level set methods allow user interactive (real-time) segmentation because the level set computation and graphic rendering are simultaneously handled in the GPU memory.

*A parallel algorithm is work-efficient, or simply "efficient", if it performs the same amount of work, to within a constant factor, as the fastest known sequential algorithm. And a parallel algorithm is step-efficient if the amount of steps is comparable to known sequential algorithms. For example, we want to add a constant value to each element of an N elements array. There will be N addition operations (work) to achieve the task for both sequential and parallel algorithms. However if we have N processors from our GPU, the parallel algorithm could map each addition operation on one processor and finish the whole operations within one step (N operations / N processor). So the work complexity of the parallel algorithm is N, while its step complexity is 1.

Chapter 3

INTELLIGENT INITIALIZATION AND INTERACTIVITY: OPTIMIZING LEVEL SETS FOR T1-WEIGHTED WHITE MATTER SEGMENTATION

The level set surface evolution is controlled by the speed function (Eq. 2.2), which requires users to specify serveral parameters, which are related to local intensities and growth penalties based on surface smoothness. Each parameter demands user experience with the tool, intuition relating parameter adjustments to changes in the result, and knowledge of the intensities and surface smoothness of the target brain structures. First, we present a method to optimally specify the three parameters to segment region of interest (ROI) from medical images directly without user input. We develop a novel non-parametric data term and optimally specify α based on imaging noise levels. In addition, we demonstrate the user interactive segmentation tool. It has a graphical user interface and could visualize the medical images and segmentation in 2D and 3D view. This method was evaluate to segment brain white matter from T1-weighted MRI phantom data, of which ground truth is provided.

My contributions in this manuscript are 1) algorithm development and implementation, 2) method evaluation, and 3) drafting and revising the manuscript.

This manuscript is presented in the Interactive Medical Image Computing (IMIC) Workshop at MICCAI 2014 conference.

Authors: Wenzhe Xue[1], Christine M. Zwart[2], and J. Ross Mitchell[3]

Affiliation:

1. Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ
2. Department of Radiology, Mayo Clinic, Scottsdale, AZ

3. Department of Research, Mayo Clinic, Scottsdale, AZ

Abstract White matter (WM) segmentation from T1-weighted MRI is complicated by intensity non-uniformities, noise, and WM's high surface area to volume ratio. Accurate algorithms are often computationally intensive and time consuming, precluding interactivity and routine clinical use. To address this we developed a work- and step-efficient parallel narrow-band level set algorithm and mapped this onto commodity GPU hardware. Our algorithm can segment brain WM in 3 seconds. However, it requires expert tuning of 3 parameters. Here we describe recent efforts to improve the precision, accuracy and simplicity of WM segmentation by: a) intelligently initializing algorithm parameters; and, b) allowing interactive parameter tuning during algorithm execution, along with real-time 2D and 3D visualization of parameter effects on segmentation results.

3.1 Introduction

MRI provides high spatial resolution and soft tissue contrast. Consequently, it is widely used for research and clinical neurological studies. Changes in brain morphometry have been linked to childhood development, healthy aging, neurological disorders, and psychiatric disorders. In particular, white matter (WM) develops dramatically through infancy (Giedd *et al.*, 1999) and undergoes significant atrophy as part of healthy aging (Ge *et al.*, 2002). WM volume changes can be a very sensitive (if not specific) indicator of Alzheimer's disease and other disorders (Hirono *et al.*, 2000; Sanfilipo *et al.*, 2006).

In the past two decades, a wealth of MR image segmentation algorithms have been developed for automatic or semi-automatic segmentation of WM and other brain

structures. Segmentation techniques range from intensity thresholding, to region-based or edge-based methods, to active contour models and level set methods. Despite advances in theoretical approaches, clinics and clinical research laboratories still rely heavily on trained technicians manually delineating regions of interest on each 2D cross-section and then extrapolating to 3D surfaces (Brinkley and Rosse, 2002). The computation time required for many sophisticated algorithms precludes routine clinical use.

Minimally interactive, user-initiated approaches with manual refinement are often proposed as an ideal solution (Brinkley and Rosse, 2002). Among these algorithms, the level set approach is popular because of its flexibility and robustness (Sethian, 1999; Osher and Fedkiw, 2001). The level set (and most region growing) approaches begin with a user-placed seed region of interest (ROI), for example in the WM on the image. The algorithm then iteratively deforms the seed ROI to encompass the entire 'related' area, in our scenario, the entire WM. In addition to the seed region, several algorithm parameters explicitly and intrinsically control the growth and smoothness of the ROI surface. These may include the relative weights of external forces acting on the contour, intensity threshold(s), and a surface smoothness parameter.

Even when the user specification of parameters is well-managed and streamlined (such as in ITK-SNAP (Yushkevich *et al.*, 2006), a popular semi-automatic segmentation tool), long computation times preclude a truly 'interactive' approach to segmentation. This is particularly true for level sets - the delay between parameter adjustments and outcomes prohibits fine tuning. Approaches to streamline parameter adjustment into a sequential process provides a false sense of fine tuning as the coarse parameters become unavailable baring a fresh start.

A recent work- and step-efficient GPU implementation of the level set method

(Roberts *et al.*, 2010) has provided a 14x acceleration over the best previously published GPU approach (Lefohn *et al.*, 2003a). A validation study of brain tumor segmentation with the new algorithm showed tumor segmentation times averaging just 1:20 (mm:ss). This was slightly *faster* than diameter-based approaches frequently used in clinical trials, yet provided volume quantification results statistically equivalent to a consensus of manual, slice-by-slice, contour delineations by experts (gold standard) (Dang *et al.*, 2013). Furthermore, the speed function used in those studies (Roberts *et al.*, 2010; Dang *et al.*, 2013) to control the level set surface evolution required specifying only three parameters (intensity window and level and a growth penalty based on surface smoothness). Nevertheless, tuning each parameter to obtain acceptable results required user experience and intuition. Interactive parameter adjustment during algorithm execution, coupled with real time visual feedback of parameter effects on segmentation results, can significantly narrow the intuition gap between novice and experienced users. This is particularly true for high contrast spherical structures (e.g., meningiomas). MRI intensity inhomogeneity, noise, and the complicated surface structure of WM hinders effective parameter selection and, in turn, rapid and accurate brain WM segmentation.

In this paper, we describe a novel approach to initializing the three parameters for GPU level set segmentation of WM in T1 weighted MRI brain images, directly, without user input of specific thresholds or values. Initially, users mark foreground and background seed pixels using a paint-brush style interface. A fuzzy classification model of the seed ROI intensities then controls the local growth and contraction of the level set. Next, we use an empirically determined relationship between curvature, segmentation accuracy, and image noise to predict the optimal curvature influence. Finally, the graphical user interface permits visualizing the growth of the level set

curve as it is computed and permits adjustment of the curvature parameter in real time while observing the surface progression in two or three dimensions. Each of these aspects is reviewed in the Methods (Section 2) prior to coverage of our white matter segmentation results (Section 3) and a brief discussion (Section 4). Videos of our tool and other additional materials have been provided online as well (see Section 5).

3.2 Methods

3.2.1 Level Sets

Level set methods embed an implicit surface within an image, and iteratively deform the surface to envelop the ROI. Several comprehensive reviews of level set methods and their application to image segmentation are well-documented in the medical image analysis literature (Sethian, 1999). Here we review the formulation relevant to our algorithm.

The implicitly represented level set surface is defined as $\{\mathbf{x}|\phi(\mathbf{x},t)=0\}$, where $\mathbf{x}$ is a coordinate in the image volume, t is the current iteration time in the level set evolution, and $\phi(\mathbf{x},t):\Re^4\mapsto\Re$ refines the level set according to:

$$\phi(\mathbf{x},t)=\phi(\mathbf{x},t-\Delta t)+\nabla t\cdot F(\mathbf{x},t)|\nabla\phi(\mathbf{x},t-\Delta t)|. \tag{3.1}$$

A speed function $F(\mathbf{x})$ defines the rate of motion of each local point on the implicit surface. The deforming direction of each point is along the norm of the local surface. Our previous GPU level set algorithm (Roberts *et al.*, 2010) adopted the speed function proposed in (Lefohn *et al.*, 2003a):

$$F(\mathbf{x},t)=\alpha C(\mathbf{x},t)+(1-\alpha)D(\mathbf{x}), \tag{3.2}$$

where $C(\mathbf{x}, t)$ is the curvature term, $D(\mathbf{x})$ is the data term, and $\alpha \in [0, 1]$, the weight of curvature term, is a blending term that controls the relative contribution of the curvature and data terms.

The data term used in (Lefohn *et al.*, 2003a) is a function of intensities in a single image volume:

$$D(\mathbf{x}) = \epsilon - |I(\mathbf{x}) - T|, \tag{3.3}$$

where $I(\mathbf{x})$ is the image intensity at location $\mathbf{x}$, T is a user-specified target intensity that encourages maximum level set growth, and ϵ is a user-specified parameter that indicates the range of intensities around T that will promote level set growth. If $T - \epsilon < I(\mathbf{x}) < T + \epsilon$, then $D(\mathbf{x})$ will promote surface growth. Otherwise it will promote surface contraction.

The curvature term $C(\mathbf{x}, t)$ depends upon the mean curvature of the local surface from the previous iteration:

$$C(\mathbf{x}, t) = \nabla \cdot \frac{\nabla \phi(\mathbf{x}, t - \Delta t)}{|\nabla \phi(\mathbf{x}, t - \Delta t)|}. \tag{3.4}$$

Penalizing curvature prevents critical leakage through weak boundaries and fills holes caused by noise inside the segmented region. Excessive smoothing by overly penalizing curvature (assigning larger value to α) can significantly distort the shape of target objects. Eventually, when α is too large (close to 1), the level set surface will shrink to a point.

The speed function defined by Eq. 3.2 relies on three user-specified parameters: α, ϵ, and T. T and ϵ define the local intensity force; α defines the contribution of the curvature term. In this paper, we describe a non-parametric approach for specifying the data term, replacing T and ϵ with user seeding of foreground and background pixels. We further suggest an approach to initializing α based on the application (WM

segmentation in T1 weighted MR) and image noise, thereby minimizing non-intuitive user input.

3.2.2 Non-parametric Data Term

We propose a novel non-parametric data term (in place of Eq. 3.3) based on the k nearest neighbor (k-NN) algorithm (Duda *et al.*, 2000). k-NN algorithms classify objects by assigning the label of majority among the k nearest training samples in feature space. The training data contain samples belonging to either foreground or background. This introduces the added requirement for users to seed both foreground and background voxels. The data term is then defined as a function of relative distance (intensity difference) to the background and foreground classes.

For voxel $\mathbf{x}$ with intensity $I(\mathbf{x})$, the distance to background $d_B(\mathbf{x})$ and distance to foreground $d_F(\mathbf{x})$ samples are assigned as the mean of distances to the k_B and k_F nearest background and foreground samples:

$$d_B(\mathbf{x}) = \frac{1}{k_B} \sum_{k=1}^{k_B} |I(\mathbf{x}) - V_B(k)|; \tag{3.5}$$

$$d_F(\mathbf{x}) = \frac{1}{k_F} \sum_{k=1}^{k_F} |I(\mathbf{x}) - V_F(k)|. \tag{3.6}$$

$k_B = \sqrt{N_B}$ and $k_F = \sqrt{N_F}$, where N_B and N_F are the numbers of background and foreground samples respectively. V_B and V_F are sorted feature (intensity) vectors of background and foreground samples.

The new data term is formed based on both distances:

$$D(\mathbf{x}) = \frac{d_B(\mathbf{x}) - d_F(\mathbf{x})}{d_B(\mathbf{x}) + d_F(\mathbf{x})}, \tag{3.7}$$

and is in the range $[-1, 1]$. Positive values indicate the voxel is closer to the foreground class, and negative values indicate the voxel is closer to the background class. The

function is zero-valued when the distances to both classes are equal. Substituting Eq. 3.7 for Eq. 3.3 eliminates user-defined parameters T and ϵ.

3.2.3 Curvature Weighting Term

In Eq. 3.2, the user-controlled weighting parameter α determines the relative cost of increased curvature. Changes in α over the full theoretical range of values ($\alpha \in (0, 1)$) can significantly impact the segmentation results. For a given application the reasonable range of values for α can be greatly reduced and the initial value for α can be automatically assigned more intelligently than, for example, setting an arbitrary starting point (e.g., $\alpha = 0.5$). This places the user into a fine-tuning and final adjustment role that requires less time, training, and improves the overall algorithm precision. In Section 3.3.2, we briefly review experiments used to define the optimal starting value for α for based on an empirical relationship between curvature, image noise, and WM segmentation accuracy (Figure 9).

3.2.4 Software and User Interface

Our tool provides an interactive rendering window with 2D axial or 3D views. Users can quickly switch between 2D and 3D view modes with a shortcut key. In the 3D view, our tool renders the image volume as well as the segmentation results, allowing qualitative validation.

In the first step of the segmentation, users initialize the level set by sketching seed points on axial slices with a paint-brush tool. Users are required to label background samples (out of the targeted ROI) and foreground samples (in the targeted ROI). After

pressing the "play" button, the level set iteratively grows to envelop the structures of interest. Due to our unique GPU implementation, both the level set and visualization algorithms share data buffers. This allows interactive tuning of algorithm parameters along with real-time 2D or 3D visualization of the parameter effects on segmentation results. Users can go back to the previous segmentation step using the undo button. In addition, users can manually edit the segmentation results (adding or erasing) using a paint-brush tool with an adjustable width parameter in the 2D axial slices.

3.3 Results

3.3.1 White Matter Segmentation

We performed WM segmentation on synthetic MRI brain phantoms generated from the BrainWeb Simulated Brain Database (Collins *et al.*, 1998). The BrainWeb phantoms are available with a variety of realistic noise characteristics and the ground truth classification of each voxel is known. We tested on six noise levels (SNR = ∞, 100, 33, 20, 14, 11) with 0% RF inhomogeneity, used eight sets of seed points, and sixteen α values ranging from 0 to 0.30 (step-size of 0.02). RF inhomogeneity was not varied due to the availability of effective approaches to bias correction (Tsang *et al.*, 2008). The eight sets of seed points were placed by 3 independent users and saved for repeated measurements. In total 768 WM segmentations were performed (6 SNR levels x 8 sets of seed regions x 16 α values). For each segmentation, accuracy was evaluated relative to the BrainWeb classification (truth) by computing Dice's

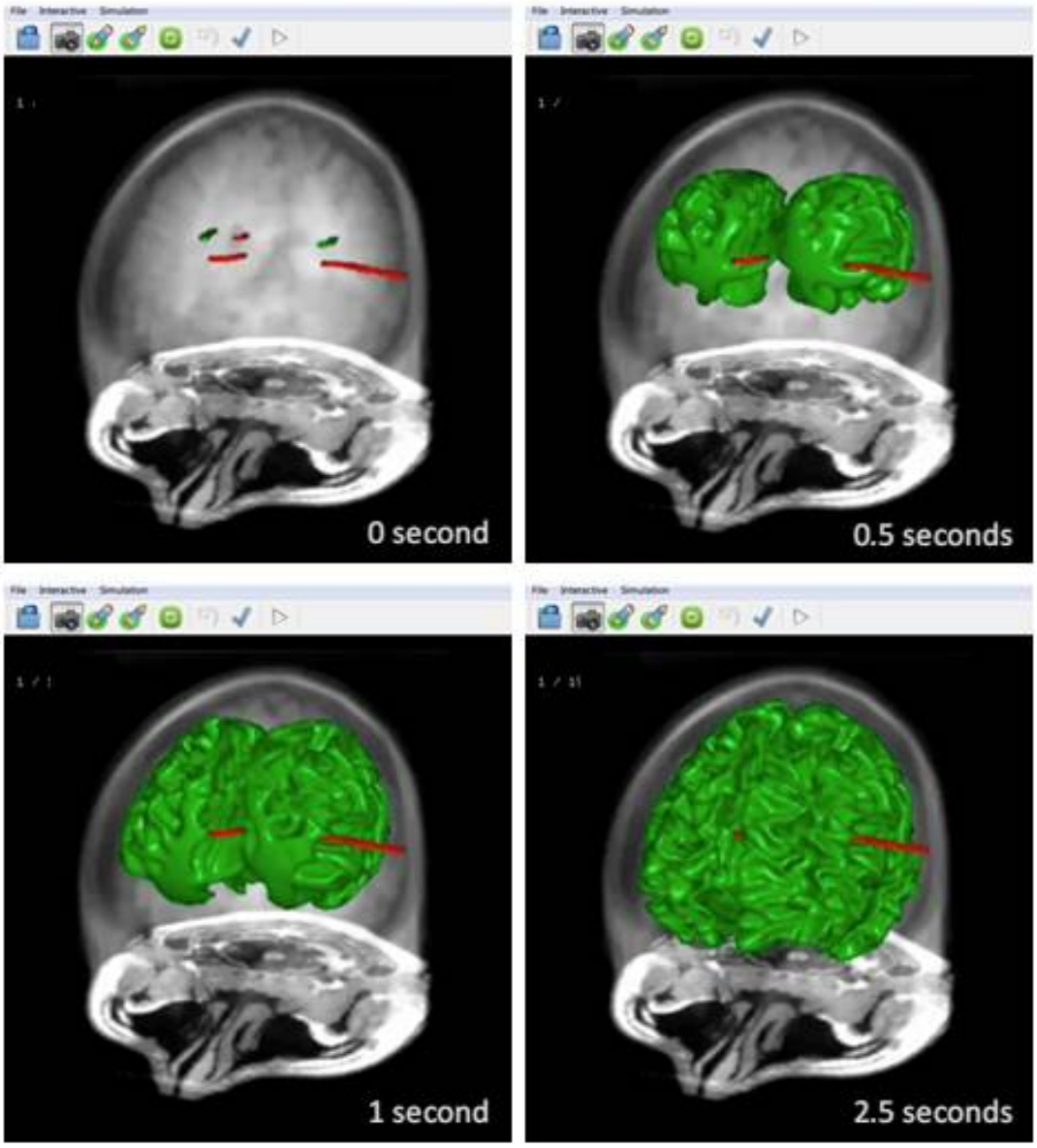

Figure 6. The 3D view of level set propogation in BrainWeb data (SNR = 33 and RF = 0%). The overall accuracy (Dice coefficient) of level set segmentation of white matter across the entire brain was 97%. The GPU level set (including kNN computation) required 3s to execute. Total segmentation time, inlcuding data loading and user interaction, was 20s.

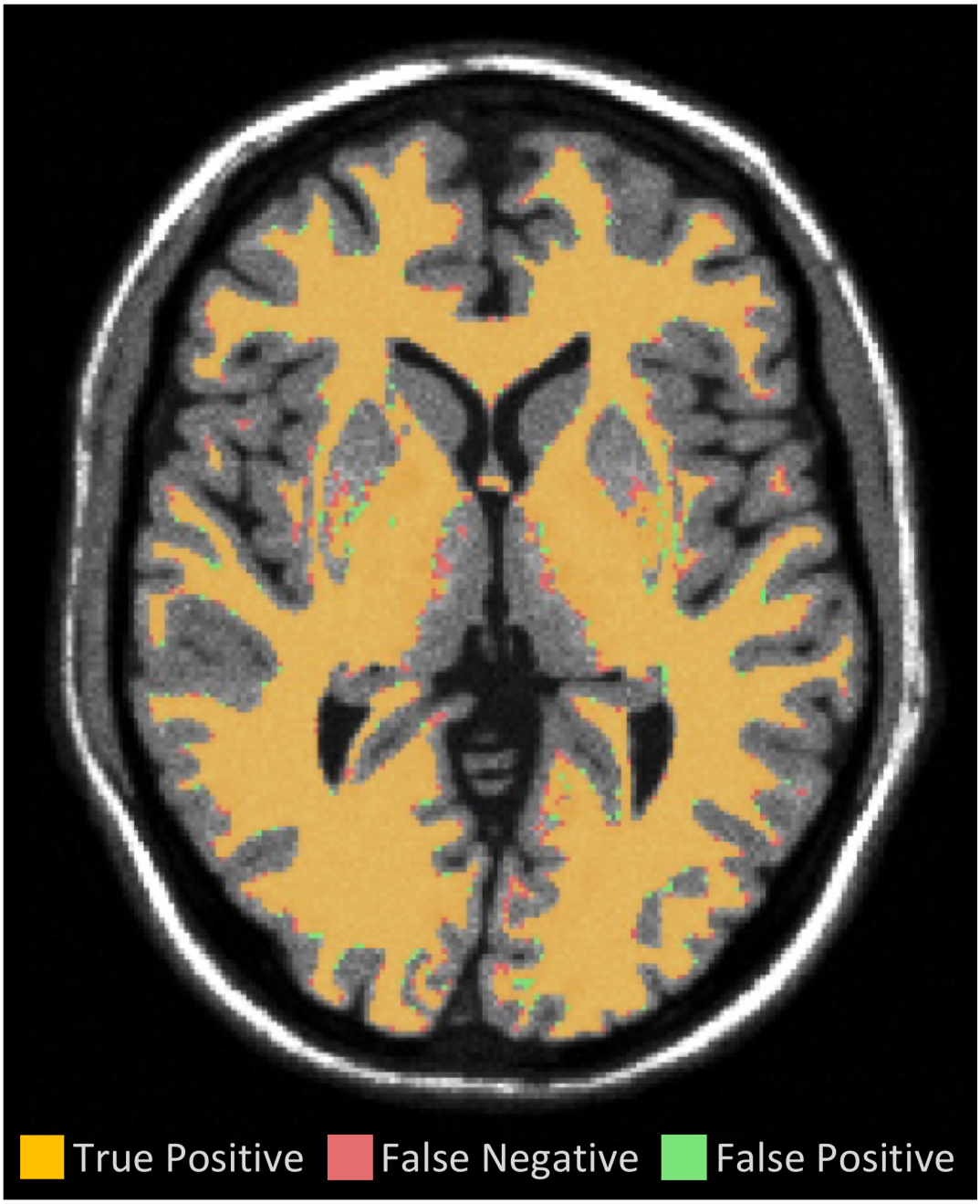

Figure 7. The difference between our level set segmentation and Brainweb ground truth, in the axial slice with the greatest number of errors.

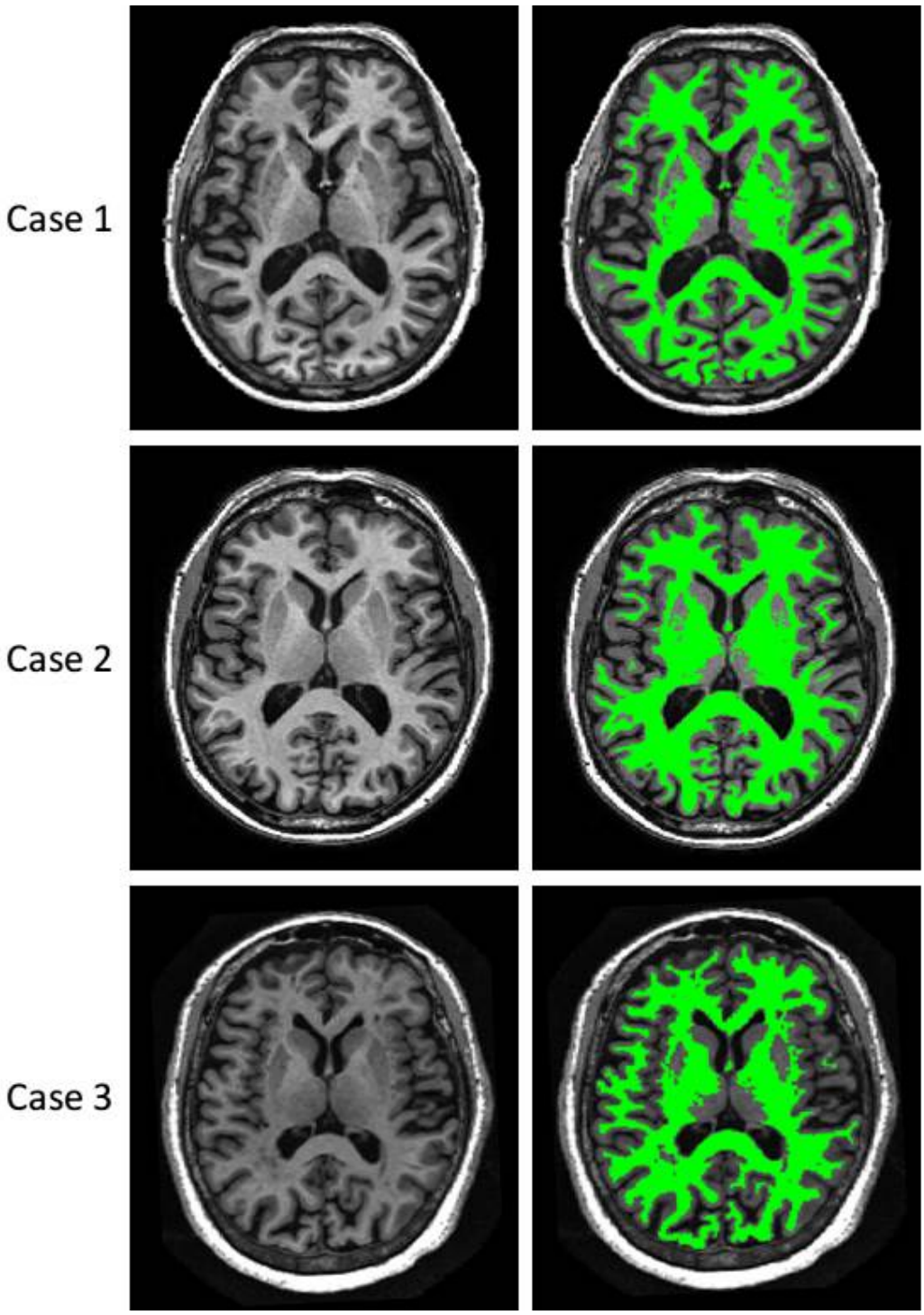

Figure 8. Level set results on patient images. The figures in the left column show the axial slices (selected to be at a location similar to Fig. 7. The right column shows the segmentation results obtained using our method.

coefficient (Dice), a measure of spatial overlap ranging from 0 to 1, with 1 indicating perfect overlap.

3.3.2 Optimized Curvature Term

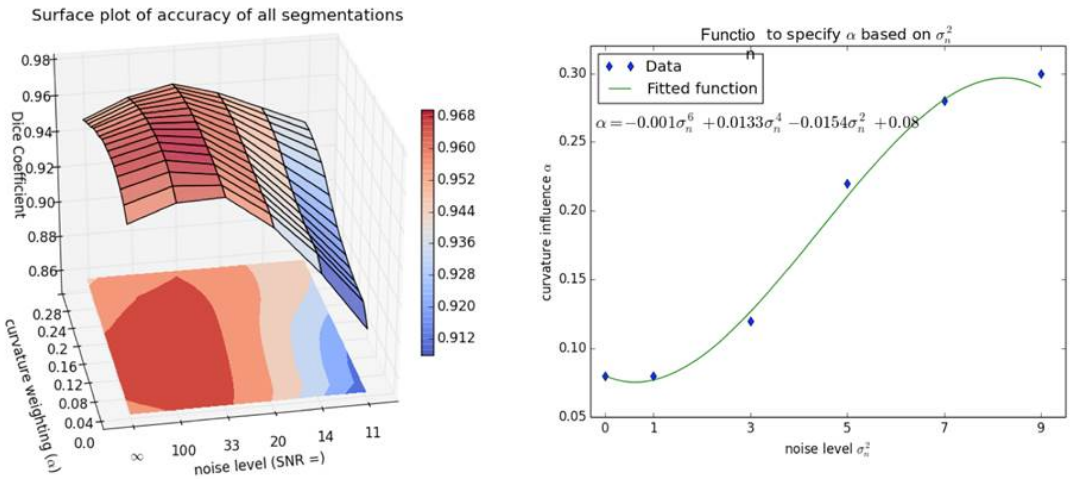

Figure 9. The relationship between accuracy, SNR, and α for 768 WM segmentations of BrainWeb T1-weighted MR phantoms. A surface plot (left), shows the mean Dice's coefficient (accuracy) (z-axis) for different SNR values (x-axis) and curvature parameters (y-axis). We then fit a function along the path of peak accuracy in this surface. The resulting 3rd order polynomial (right) allows us to estimate the α value that maximizes accuracy for a particular image noise level σ_n^2.

Plotting the Dice's coefficient for each of the segmentation experiments described above allowed us to establish an empirical relationship between the image noise, α, and accuracy (Figure 9). Using this relationship, the initial curvature weight, α_{init}, can be selected (based on estimates of image noise) according to a third degree polynomial (determined by a curve fitting algorithm):

$$\alpha_{init} = f(\sigma_n^2) = -0.001(\sigma_n^2)^3 + 0.0133(\sigma_n^2)^2 - 0.0154\sigma_n^2 + 0.08 \tag{3.8}$$

For clinical data with unknown noise, the estimated noise variance, $\sigma_n^2 = 1/\text{SNR}$,

can be computed using the fast noise estimation method proposed by Immerker (Immerkæ r, 1996). The curvature parameter can then be initialized using Eq. 3.8 without any user experience or intuition (users are still able to tune α afterward).

Over the range of image noise levels evaluated, the value for α (taken from the discrete subset tested) corresponding to the average (over the eight seed point sets) of the peak Dice coefficients varied considerably (Table 2). In contrast, the optimal α for each seed point set in a fixed noise level varied much less and the fine-tuning of α from the noise-based initial value to the experimentally-defined optimal α had limited impact on the accuracy (Table 3).

Table 2. The maximum Dice and corresponding α for each SNR value tested. The optimal α values range from 0.08 to 0.3.

SNR	100	33	20	14	11
Max. Dice	0.9712	0.9674	0.9550	0.9419	0.9310
α	0.08	0.12	0.22	0.28	0.3

Table 3. The maximum Dice and corresponding α for each set of input seed points when SNR=33. For comparison, the accuracy for fixed $\alpha = 0.12$ is also provided.

User	1	2	3	4	5	6	7	8
Max. Dice	0.9669	0.9683	0.9682	0.9679	0.9683	0.9671	0.9664	0.9657
α	0.06	0.10	0.12	0.12	0.12	0.14	0.12	0.14
Dice ($\alpha = 0.12$)	0.9668	0.9683	0.9682	0.9679	0.9683	0.9671	0.9664	0.9656

3.4 Discussion

Requiring users to adjust parameters increases segmentation variability, and inconvenience. Iterative refinement of parameters may also increase segmentation time. In general, the tuning process requires expertise and hinders use in clinical settings.

Our new approach provides precise, accurate, rapid and simple WM segmentation in T1-weighted brain MR scans. It does not require prior knowledge of appropriate parameter values, nor extensive parameter tuning to obtain acceptable results. We accomplished this through three novel contributions: 1) a GPU level set speed function driven by a non-parametric k-NN data model built from the seed ROI intensities; 2) a new empirical relationship between WM segmentation accuracy, image noise, and the level set curvature parameter, α; and, 3) a GPU implementation where both the level set and visualization algorithms share data buffers. This allows interactive tuning of algorithm parameters along with real-time 2D or 3D visualization of the parameter effects on segmentation results.

Future efforts will focus on validating our method for WM segmentation in other imaging modalities, such as T2-weighted MRI. We will also extend the method for segmentation of brain gray matter and ventricles.

3.5 Supplementary Material

We have used our tool to segment a variety of objects from MR and CT datasets. A video of our tool segmenting brain WM (from T1-weighted MRI) and brain vasculature (from contrast enhanced CT) is available here: http://goo.gl/2K98Gs.

Chapter 4

A NEW ITERATIVE GPU ALGORITHM TO SEGMENT MS LESIONS IN MULTI-SPECTRAL MRI DATASETS

As mentioned in Chapter 2, multiple medical imaging modalities add much more information to monitor the disease progress during treatment. In this chapter, we present an novel method to extract brain lesions from multiple segmentations of brain white matter from multi-spectral MRI. MS lesion is very difficult to identify and extract from a single contrast MRI. It's necessary to use multi-spectral images. We explore a unique approach of using our previous algorithm to solve such a sophisticated task. In particular, this new approach is feasible only by leveraging the speed of our GPU level set algorithm. We evaluate our method on both MRI phantom and real patient MRI data. We demonstrate that our method is able to achieve the accuracy comparable to the state-of-the-art algorithms in seconds.

My contributions in this manuscript are 1) algorithm development and implementation, 2) method evaluation, and 3) drafting the manuscript.

Currently, we plan to submit this manuscript to the journal of NeuroImage.

Authors: Wenzhe Xue[1], Christine M. Zwart[2], Joseph M. Hoxworth[2], Dean M. Wingerchuk[3], and J. Ross Mitchell[4]

Affiliation:

1. Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ
2. Department of Radiology, Mayo Clinic, Scottsdale, AZ
3. Department of Neurology, Mayo Clinic, Scottsdale, AZ
4. Department of Research, Mayo Clinic, Scottsdale, AZ

Abstract This paper presents a new method to segment MS lesions in multi-spectral MRI exams. Our approach leverages the speed of a GPU level set algorithm to perform multiple segmentations of brain white matter in each exam. Differences between these segmentations allow us to estimate the distribution of MS lesions. A new data term in the level set speed function eliminates the need to store MR exams on the GPU memory. This allows multi-spectral MRI exams that are too large to fit on the GPU to be processed efficiently. We evaluated our method on BrainWeb and MS lesion segmentation challenge 2008 data. It achieved results comparable to, or better than, two top performing algorithms. The average time required by our algorithm to segment lesions in the MS Lesion segmentation Challenge 2008 data was 20.8 seconds. Current limitations include the need for a user to seed three regions in each exam. Future work will focus on automating the seed placement step using statistical and/or atlas-based co-registration approaches. Finally, better post-processing methods may reduce false positive lesions.

4.1 Introduction

Multiple sclerosis (MS) is a chronic inflammatory demyelinating disease of the central nervous system (CNS). Typical manifestations of MS include motor weakness, gait disorder, and impairment of sensory, visual, and cognitive function. These are caused, in part, by focal inflammatory lesions that disrupt CNS white matter pathways. Magnetic resonance imaging (MRI) is widely used to diagnose MS and evaluate disease activity over time.

Focal MS white matter lesions typically appear hyperintense on T2 weighted (T2w) and fluid-attenuated inversion recovery (FLAIR) T2 MRIs, and isointense or

hypointense on non-contrast enhanced T1 weighted (T1w) MRIs. Such lesions can be visually distinguished from surrounding "normal-appearing white matter" (NAWM).

Accurate and reproducible quantitative measurement of MS lesion volumes aids disease tracking, therapeutic assessment, and cognitive research. However, segmentation of MS lesions is generally quite difficult due to their highly variable size, distribution, and intensity patterns. Consequently, manual identification and segmentation of MS lesions is often performed by experts such as physicians or trained technicians. But this process is time consuming and plagued by low intra- and inter-operator reliability.

Completely automated approaches are also used. These often employ two important components to extract MS lesions: preprocessing and segmentation. Preprocessing of head MRI typically includes spatial co-registration, intensity inhomogeneity correction, and brain extraction. Intensity normalization is used when the segmentation component involves a training process and thus requires the intensity of the analyzed image to be similar to the intensity of training images. Image features used for automatic identification of MS lesions include normalized intensity, texture measures, location and shape (García-Lorenzo *et al.*, 2013). Completely automated approaches may require long execution times. Furthermore, in clinical research settings, their output must be reviewed by a human expert for both false-positive and false-negative results. The review and correction process can be time consuming.

Minimally interactive, user-initiated approaches with manual refinement are often proposed as an ideal compromise for general medical image segmentation (Brinkley and Rosse, 2002; Menze *et al.*, 2014). These methods combine the human expert's ability to identify normal and pathological regions, with the computer's computational power. Computer assisted MS lesion detection and segmentation methods have been

an active area of research for several decades (Brosch *et al.*, 2015; Freire and Ferrari, 2016; García-Lorenzo *et al.*, 2013, 2009; Mitchell *et al.*, 1994; Styner *et al.*, 2008).

This paper describes two primary contributions. The first of these is a novel method to segment MS lesions from multi-spectral MRI. Our approach builds upon an existing general purpose GPU level set algorithm (Roberts *et al.*, 2015, 2010). This GPU algorithm is optimal in the sense that it is both work and step efficient. It is also 14x faster than the fastest prior parallel level set algorithm (Lefohn *et al.*, 2004). It allows rapid, low bias and low variability segmentation of neurological tissues, including brain tumors (Dang *et al.*, 2013) and total white matter (WM) volume (Xue *et al.*, 2014). The latter required less than 3 seconds to complete.

The second primary contribution is a new data term in the level set speed function. This term eliminates the need to store MRI exams on the GPU memory, as in prior approaches. Instead, a fixed size data and seed dependent lookup table is employed. This allows multi-spectral MRI exams that are too large to fit on the GPU memory to be processed efficiently using our algorithm.

4.2 Methods

4.2.1 Development and Validation Environment

All development and validation were performed on a workstation running Windows 7 (Microsoft, Redmond WA.). This workstation had an Intel (Santa Clara, CA.) Xeon 8-core 3.6GHz CPU, 64GB RAM, and an Nvidia (Santa Clara, CA.) GTX Titan GPU (6GB VRAM and 2688 CUDA cores). Our application was implemented in C++ and CUDA (V6.0) using Visual Studio 2008 (Microsoft).

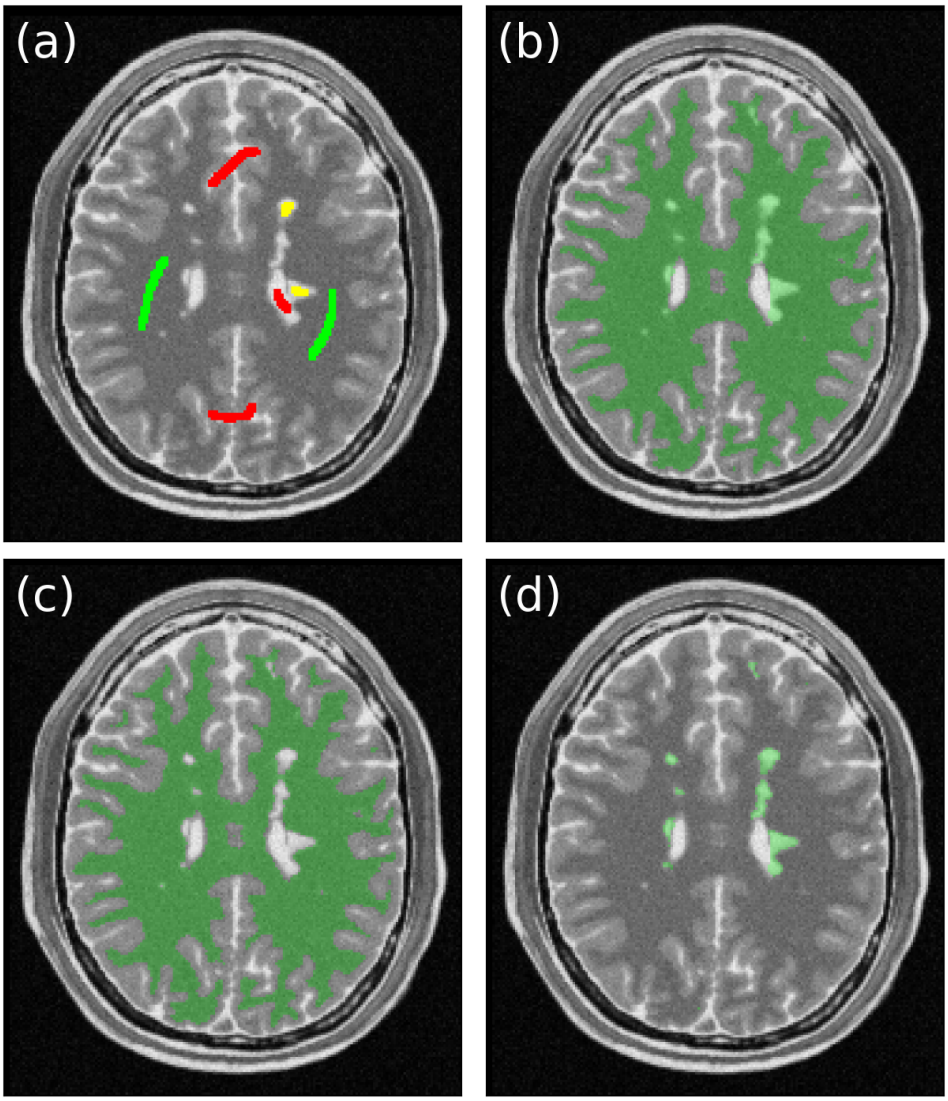

Figure 10. MS lesion segmentation on a representative axial T2-weighted image from the BrainWeb data with moderate lesion load. Segmentation proceeds in four steps: (a) the user places seeds: NAWM seeds are green; CSF and gray matter (GM) seeds are red; and, lesion seeds are yellow. (b) the GPU level set labels all white matter (AWM, including NAWM and MS lesions) in 3D; (c) the GPU level set labels NAWM in 3D; (d) the post-processed difference between (b) and (c) is used to estimate the distribution of MS lesions. A close up comparison between our estimate and truth is shown in Fig. 11. When our method was tested on the MS Lesion Segmentation Challenge 2008 dataset, the average execution time per exam was 20.8 seconds.

4.2.2 Background

4.2.2.1 Level Set Methods

Level set methods embed an implicit surface within an image, and iteratively deform the surface for segmentation (Sethian, 1999). The implicitly represented level set surface is defined as $\{\mathbf{x}|\phi(\mathbf{x},t)=0\}$, where $\mathbf{x}$ is a coordinate in the image volume, t is the current iteration time in the level set evolution, and $\phi(\mathbf{x},t):\Re^4\mapsto\Re$ refines the level set according to:

$$\phi(\mathbf{x},t)=\phi(\mathbf{x},t-\Delta t)+\nabla t\cdot F(\mathbf{x},t)|\nabla\phi(\mathbf{x},t-\Delta t)|. \tag{4.1}$$

A speed function $F(\mathbf{x})$ defines the rate of motion of each local point on the implicit surface. The deforming direction of each point is along the norm of the local surface. Our previous GPU level set algorithm adopted the speed function proposed in (Lefohn *et al.*, 2004):

$$F(\mathbf{x},t)=\alpha C(\mathbf{x},t)+(1-\alpha)D(\mathbf{x}), \tag{4.2}$$

The curvature term $C(\mathbf{x},t)$ depends upon the mean curvature of the local level set surface from the previous iteration:

$$C(\mathbf{x},t)=\nabla\cdot\frac{\nabla\phi(\mathbf{x},t-\Delta t)}{|\nabla\phi(\mathbf{x},t-\Delta t)|}. \tag{4.3}$$

And the data term is defined as a function of intensities in a single image volume:

$$D(\mathbf{x})=\epsilon-|I(\mathbf{x})-T|, \tag{4.4}$$

where $I(\mathbf{x})$ is the image intensity at location $\mathbf{x}$, T is a user-specified target intensity that encourages maximum level set growth, and ϵ is a user-specified parameter that

indicates the range of intensities around T that will promote level set growth. If $T - \epsilon < I(\mathbf{x}) < T + \epsilon$, then $D(\mathbf{x})$ will promote surface growth. Otherwise it will promote surface contraction.

The parameter α, in the range [0, 1], controls the relative contributions of the curvature and data terms. Our level set executes quickly enough that surface growth can be visualized interactively in two or three dimensions. Consequently, the user can adjust α and see the effect in real time. Higher α values encourage smoother, smaller volumes, while lower values encourage larger, more irregular volumes. Several adjustments to α are sometimes required to obtain a satisfactory segmentation, especially for complex shapes such as brain white matter. To address this, we previously developed an empirical approach to optimally initialize α based on image noise (Xue *et al.*, 2014). In this paper, similar experiments were performed to find optimal α values for MS lesion segmentation in patient exams (described below).

4.2.2.2 Non-parametric Data Term

In (Xue *et al.*, 2014) we proposed a non-parametric data term $D(\mathbf{x})$ based on the k nearest neighbor (kNN) classification (Duda *et al.*, 2000). This eliminated the need for the user to specify T and ϵ. Instead, the user performed a more intuitive operation: they drew one or more short strokes (placed seeds) in desired tissues (foreground, colored green in our implementation), and undesired tissues (background, colored red) (Fig. 10a).

$D(\mathbf{x})$ is then a function of the relative difference, or distance, between $I(\mathbf{x})$ and the background and foreground intensities determined from the seed samples. For a voxel $\mathbf{x}$, the distance to background, $d_B(\mathbf{x})$, and distance to foreground, $d_F(\mathbf{x})$, samples are

assigned the mean of distances to the k_B and k_F nearest background and foreground samples:

$$d_B(\mathbf{x}) = \frac{1}{k_B} \sum_{k=1}^{k_B} |I(\mathbf{x}) - V_B(k)|; \tag{4.5}$$

$$d_F(\mathbf{x}) = \frac{1}{k_F} \sum_{k=1}^{k_F} |I(\mathbf{x}) - V_F(k)|. \tag{4.6}$$

$k_B = \sqrt{N_B}$ and $k_F = \sqrt{N_F}$, where N_B and N_F are the numbers of background and foreground samples respectively. V_B and V_F are feature (intensity) vectors of background and foreground samples, sorted by distance to $I(\mathbf{x})$.

The new data term is then defined as:

$$D(\mathbf{x}) = \frac{d_B(\mathbf{x}) - d_F(\mathbf{x})}{d_B(\mathbf{x}) + d_F(\mathbf{x})}, \tag{4.7}$$

and is in the range $[-1, 1]$. Positive values promote surface growth, while negative values promote contraction. The function is zero-valued when the distances to both classes are equal.

4.2.2.3 Extension to Multi-spectral Image Volumes

We extended $I(\mathbf{x})$ to be a vector, supporting multiple contrast values at each spatial location. Equations (4.5) and (4.6) are modified to indicate vector valued image volumes:

$$d_B(\mathbf{x}) = \frac{1}{k_B} \sum_{k=1}^{k_B} |\vec{I}(\mathbf{x}) - V_B(k)|; \tag{4.8}$$

$$d_F(\mathbf{x}) = \frac{1}{k_F} \sum_{k=1}^{k_F} |\vec{I}(\mathbf{x}) - V_F(k)|. \tag{4.9}$$

The kNN distance computation to background and foreground samples then occurs in a higher-dimensional feature space. Also, since $D(\mathbf{x})$ depends only upon $d_B(\mathbf{x})$ and $d_F(\mathbf{x})$ it can be pre-computed on the CPU and transferred to the GPU. This eliminates the requirement to store $\vec{I}(\mathbf{x})$ on the GPU, as in (Roberts *et al.*, 2010, 2015). This is significant since $\vec{I}(\mathbf{x})$ may be too large to fit into GPU memory, especially for high spatial resolution image volumes with multiple intensity values at each spatial location. The drawback is that $D(\mathbf{x})$ must be recomputed on the CPU and transferred to the GPU whenever the user specifies new background or foreground seeds, since that could alter $d_B(\mathbf{x})$ or $d_F(\mathbf{x})$, respectively. However, this process requires only a few seconds, and is typically much quicker than the time the user required to specify new seeds. In our experience this relatively short delay is well tolerated by users.

4.2.3 MS Lesion Segmentation

Since the time penalty for segmentation with our GPU level set is small, multiple segmentations of large regions in a single neurological MRI dataset are possible in a clinically practical time frame. We leverage this to repeatedly segment the total WM volume, while varying the level set initialization parameters between iterations.

We begin by segmenting all white matter (AWM, including NAWM and MS lesions). Then, through successive iterations, we adjust the initialization parameters to constrain the segmentation towards NAWM alone. Differences between these repeated segmentations allow us to estimate the spatial distribution of MS lesions. In theory, the more segmentations we perform, the better our estimate will become. Here, as a proof of concept, we perform only two segmentations of WM - the minimum possible with our scheme.

To accomplish this, we introduce a new workflow that requires the user to label seed points for three tissues (Fig. 10a): 1) NAWM (green), 2) other brain regions (gray matter and CSF, red), and 3) one or more MS lesions (yellow). Using these three clusters of seed voxels, our method performs two separate segmentations. The first is to extract AWM, and the second is to extract NAWM.

Segmentation of AWM To segment AWM, our method uses seed voxels of NAWM and MS lesion as foreground samples, and seed voxels from other brain regions as background samples. Equations (4.8), (4.9) and (4.7) are calculated and used to guide the propagation of the level set surface. The result is an estimate of the AWM region, $\mathbf{X}_{AWM}$.

Segmentation of NAWM To segment NAWM, our method uses seed voxels of NAWM as foreground samples and seed voxels of MS lesions, and other brain regions (GM and CSF) as background samples. As above, equations (4.8), (4.9) and (4.7) are calculated and used to guide the propagation of the level set surface. The result is an estimate of the NAWM region, $\mathbf{X}_{NAWM}$.

Segmentation of MS lesions We estimate the MS lesion region as the difference between AWM and NAWM:

$$\mathbf{X}_{MS} = \mathbf{X}_{AWM} - \mathbf{X}_{NAWM}. \tag{4.10}$$

As would be expected with most subtractive techniques, some voxels along the boundaries between tissues may be included in the MS lesion results (false positives). These voxels normally form disconnected small objects, which are removed by a morphological opening operation with a radius of 1 voxel.

4.2.4 Algorithm Evaluation

We evaluated our method on two public, freely available MS lesion datasets: the BrainWeb synthetic phantom data (Cocosco *et al.*, 1997), and the 2008 MS lesion segmentation challenge data (Challenge08) (Styner *et al.*, 2008).

BrainWeb phantom The BrainWeb Simulated Brain Database provides MRI phantoms with three different MS lesion loads: mild, moderate, and severe. Each phantom contains T1w, T2w, and proton density weighted (PDw) spin-echo MRIs. Also included is a ground truth, voxel-by-voxel, tissue label map. In addition, different noise and inhomogeneity levels can be simulated and added to the images.

Our algorithm was applied to T1w and T2w MRIs with a noise level of 3% and an intensity inhomogeneity of 0%, without any preprocessing. RF inhomogeneity was not varied due to the availability of effective approaches for bias correction. Two sets of seed points were placed by one user to obtain the MS lesions in each mild, moderate, and severe case. The curvature tuning parameter α was set to 0.12 by default (Xue *et al.*, 2014). The segmentation accuracy was evaluated relative to the ground truth by computing the Dice similarity coefficient (DSC), a measure of spatial overlap ranging from 0 to 1, with 1 indicating perfect overlap. Then the DSCs were compared to the DSC reported by Garcia-Lorenzo et al. (García-Lorenzo *et al.*, 2009), the top performing algorithm in Challenge08.

We observed that the ground truth of BrainWeb data contains very small lesions, with just a few voxels each. Therefore, we also computed the lesion detection rate (LDR) for the BrainWeb phantoms, where LDR is the DSC based on the overlap of detected lesions rather than voxels. A lesion was considered “detected" if any of its voxels were labeled as lesion in the segmentation. Assessing our algorithm’s ability

to detect at least some portion of a lesion, in addition to all lesion voxels, separates the task of lesion identification from the task of volumetric segmentation. In effect, LDR assigns a greater penalty to missing (or erroneously adding) a small lesion than missing (or adding) a small region of pixels on an otherwise successfully identified lesion.

The LDR was computed only for lesions greater than four voxels in volume as smaller lesions were eliminated, along with some noise and artifacts, in the morphological opening step. We rationalized this decision after consulting with an MS neurologist at our institution (DW). Lesions consisting of 4 voxels or less are rarely measured in clinical practice. For example, there were no lesions this small in the expert manual labels provided with Challenge08 data.

Patient images Challenge08 (Styner *et al.*, 2008) provides 20 training cases, half from the Children's Hospital of Boston (CHB) and half from the University of North Carolina (UNC). Each case contains T1w, T2w, and FLAIR MRI. This data is provided with the following preprocessing steps performed: brain extraction, intensity inhomogeneity correction, and intensity re-scaling (to 8-bit unsigned integer). Manual labels of the MS lesions by one expert (from CHB) are provided and were used as truth in algorithm evaluations. Our algorithm was provided with the T2w and FLAIR MRI for MS lesion segmentation. Addition of the T1w images did not improve our results (data not shown).

For each case, three sets of seed points (described above) were placed by one user (WX) and saved for repeated use. We then performed a 5-fold cross validation of our algorithm as in Brosch et al. (Brosch *et al.*, 2015). Briefly, each case was randomly assigned to one of 5 groups. The first 4 groups were used to train our algorithm (details below). Then the trained algorithm was applied to the last "test" group, and

the results obtained. This process was repeated 5 times, selecting a different test group each time. At each step the segmentation accuracy was evaluated relative to the expert manual labels by computing the average values of the true positive rate (TPR), positive prediction value (PPV), and DSC across the 4 cases in the test group. Our results were compared to the results reported by Geremia et al. (Geremia *et al.*, 2011) and Brosch et al. (Brosch *et al.*, 2015). The execution time required to segment each test case was also recorded. These times were then averaged across all 5-folds of the cross validation to determine the overall mean execution time.

Algorithm training consisted of identifying optimal values for α in equation (4.2) for AWM and NAWM using a procedure described previously (Xue *et al.*, 2014). We refer to these as α_{AWM} and α_{NAWM}, respectively. Briefly, for each tissue α was varied from 0.05 to 0.45 with a step size of 0.1. The pair (α_{AWM}, α_{NAWM}) that maximized the DSC for $\mathbf{X}_{MS}$ was retained and applied to the test group of cases. Importantly, the training data provides guidance for parameter settings in this application. The user can make adjustments and the algorithm does not require a training step (unlike (Brosch *et al.*, 2015)).

4.3 Results

BrainWeb phantoms The DSC of our method is presented and compared to (García-Lorenzo *et al.*, 2009) (Table 4). Our method had a lower DSC on MRIs with mild lesion load, but a higher DSC on cases with moderate and severe lesion load. DSC and LDR for both unaltered and small-lesion (4 voxels or less) removed versions of truth are also presented (Table 5). Eliminating small lesions from the ground truth had a small positive impact on DSC, but increased the average LDR of our method from 0.70 to 0.89. In other words, our method detected 89% of the MS lesions larger than 4 voxels, on average.

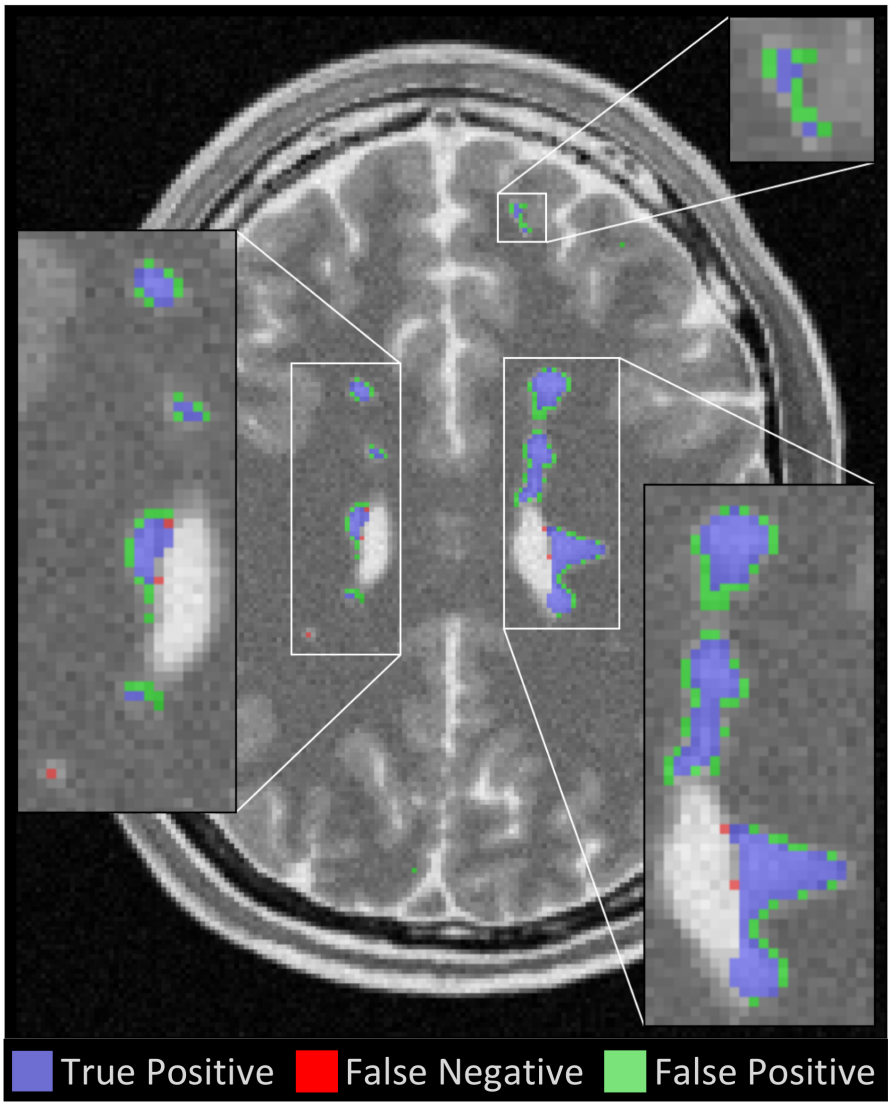

Figure 11. Close up comparison between the true MS lesion distribution and our algorithm's estimated MS lesion distribution. This is the same representative slice depicted in Fig. 10, and is from the BrainWeb phantom with moderate lesion load. True positive voxels are blue; false positive voxels are green; and, false negative voxels are red.

Table 4. Comparison of DSC on BrainWeb data

	Mild	Moderate	Severe
Garcia09	.67	.72	.76
Our method	.55	.74	.81

Patient images Mean α values and execution times from algorithm training during the 5-fold cross validation are shown in Table 6. The relationship between the pair of values (α_{AWM}, α_{NAWM}) and overall DSC for $\mathbf{X}_{MS}$ is shown in Fig. 12 for each contributing imaging center (CHB and UNC).

Results from the algorithm evaluation are shown in Table 7, and compared to published results from two leading algorithms. Our algorithm had a higher TPR and DSC than the comparison algorithms, and a PPV nearly half way between these algorithms.

Table 5. DSC and LDR (lesion detection rate) of our method on BrainWeb data

		Mild	Moderate	Severe
Based on unaltered truth	DSC	.55	.74	.81
	LDR	.67	.69	.74
Truth wo/ small lesions	DSC	.55	.75	.82
	LDR	.88	.84	.94

Table 6. Algorithm training results. α is dimensionless. Execution time is in seconds. Value shown is the overall mean ($\pm$ standard deviation) from the 5-fold cross-validation.

	α	Execution Time
AWM	0.45 ($\pm$ 0.0)	12.47 ($\pm$ 3.19)
NAWM	0.075 ($\pm$ 0.05)	8.33 ($\pm$ 4.40)
Total	–	20.80 ($\pm$ 5.43)

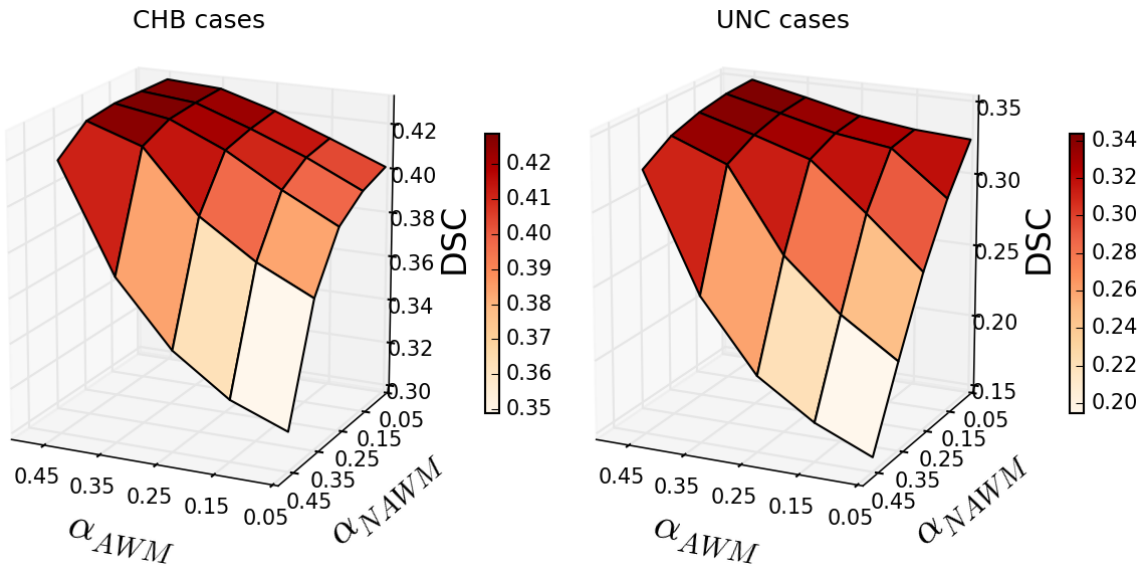

Figure 12. Graph of the dice coefficient (accuracy, vertical axis) for $\mathbf{X}_{MS}$ for various (α_{AWM}, α_{NAWM}) combinations. High values for α_{AWM} and low values for α_{NAWM} produced the highest accuracy. This observation applied equally to the data from both centers.

Table 7. Comparison of the results with other methods.

Method	TPR	PPV	DSC
Geremia et al. (Geremia *et al.*, 2011)	39.85	40.35	–
Brosch et al. (Brosch *et al.*, 2015)	39.71	41.38	35.52
Our method	44.33	40.79	38.88

4.4 Conclusion

We developed a new method to segment MS lesions in multi-spectral MRI exams. Our approach leverages the speed of a GPU level set algorithm to perform multiple segmentations of brain white matter in each MRI exam. Changes in level set initialization parameters between iterations allow us to estimate the distribution of white matter lesions. A new data term in the level set speed function eliminates the need to store MRI exams on the GPU memory. Instead, a fixed size data and seed dependent

lookup table is employed. This allows multi-spectral MRI exams that are too large to fit on the GPU memory to be processed efficiently using our algorithm.

The proposed method was evaluated on BrainWeb and Challenge08 data and achieved results comparable to, or better than, two leading algorithms. The average execution time for our algorithm to segment lesions in Challenge08 data was 20.8 seconds. Current limitations include the need for a user to seed three regions in each exam: NAWM; GM and CSF (major brain anatomy); and, one or more MS lesions. Future work will focus on automating the seed placement step using statistical and/or atlas-based co-registration approaches. Finally, better post-processing methods may reduce false positive lesions.

Chapter 5

RETROSPECTIVE VALIDATION OF A COMPUTER ASSISTED QUANTIFICATION MODEL OF INTRACEREBRAL HEMORRHAGE VOLUME, COMPARED TO ABC/2 METHOD

In this chapter, we report a retrospective clinical validation study to evaluate our method on a larger clinical dataset. This study presents a general framework to comprehensively validate medical image segmentation tools in clinical setting. Four clinicians performed intracerebral hemorrhage volume measurements. The bias, precision, and measurement time were calculated based on the results, and compared to ABC/2 method (the common clinical practice). Note that in this manuscript, we name our application SegTool, which refers to the user interactive application with graphical user interface.

My contributions in this manuscript are 1) making the changes of underlying algorithm and front-end graphical user interface of our application for ICH study, 2) training all operators, 3) collecting and cleaning the volume data after the measurements 4) drafting abstract, methods, and results and creating three of the four figures. Note that Dr Vegunta is co-first author, contributing on the clinical side of this study.

This manuscript is submitted to American Journal of Neuroradiology.

Authors: Wenzhe Xue[1], Sravanthi Vegunta[2], Christine M. Zwart[3], Maria I. Aguilar[4], Ameet C. Patel[3], Joseph M. Hoxworth[3], Bart M. Demaerschalk[4], and J. Ross Mitchell[5]

Affiliation:

1. Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ

2. Moran Eye Center, University of Utah, Salt Lake City, UT
3. Department of Radiology, Mayo Clinic, Scottsdale, AZ
4. Department of Neurology, Mayo Clinic, Scottsdale, AZ
5. Department of Research, Mayo Clinic, Scottsdale, AZ

Abstract *Background and Purpose:* Intracerebral hemorrhage accounts for 6.5% to 19.6% of all acute strokes. Initial hematoma volume and expansion are both independent predictors of clinical outcomes and mortality. Therefore, a rapid, unbiased and precise measurement of hematoma volume is a key component of clinical management. The most commonly used method is ABC/2, which results in overestimation. We developed a custom user interactive segmentation program, SegTool, using a novel GPU level set algorithm. Until now the speed, bias, and precision of SegTool had not been validated. *Methods:* In a single stroke academic center, two vascular neurologists and two neuro-radiologists independently performed a test-retest experiment that involved repeated measurements of static, unchanging ICH volumes on CT on 76 ICH cases. Measurements were made using both SegTool and the ABC/2. True ICH volumes were estimated using a consensus of repeated manual tracings by two operators. These data allowed us to estimate measurement bias, precision and speed. *Results:* The measurements with SegTool were not significantly different compared to the true ICH volumes, while ABC/2 overestimated volume by 45%. The inter-rater measurement variability with SegTool was 50% less than with ABC/2. The average measurement times for ABC/2 and SegTool were 35.7s and 44.6s, respectively. *Conclusion:* SegTool appears to have attributes superior to ABC/2 in terms of accuracy and interrater reliability without a significant delay in acquisition time, hence it could be useful in clinical practice and clinical trials.

Keyword: intracerebral hemorrhage, stroke, image quantitative, CT, radiology

5.1 Introduction

Spontaneous, nontraumatic intracerebral hemorrhage (ICH) implies bleeding into the brain parenchyma that may extend into the ventricular system and/or subarachnoid space. ICH is the cause of 6.5%-19.6% of all strokes (Feigin *et al.*, 2000). ICH has an estimated annual incidence of 12-15 per 100,000 in the USA (Woo *et al.*, 1990) and 15.9 per 100,000 worldwide (Sacco *et al.*, 2009). After ICH, active bleeding may continue for hours (T *et al.*, 1997). Mortality has been reported to be approximately 31%-34% at 7 days and 53%-59% at 1 year (Flaherty *et al.*, 2006).

Signs and symptoms of ICH are nonspecific; therefore, neuroimaging is mandatory for diagnosis (Goldstein, 2005). Computed tomography (CT) is sensitive for identifying acute hemorrhage and is considered the gold standard (Morgenstern *et al.*, 2010). 28%-38% of patients undergoing head CT within 3 hours of ICH onset have hematoma expansion of greater than one third on follow-up CT (T *et al.*, 1997). Initial hematoma volume and rate of growth are both independent predictors of clinical outcomes and mortality (Davis *et al.*, 2006). Attenuation of growth is an important treatment strategy (Davis *et al.*, 2006). Consequently, a rapid, unbiased and precise measurement of hematoma volume is an important component of clinical management.

There are a variety of methods to measure the volume of a hematoma. Common methods of lesion volume measurement are equations for a sphere, ellipsoid, or rectangular piped. The equations are applied to a lesion according to the reader's interpretation of the lesion shape. The most commonly used equation is the simplified ellipsoid formula (ABC/2) (Divani *et al.*, 2011), "where A is the greatest hemorrhage diameter by CT, B is the diameter 90 degrees to A, and C is the approximate number of CT slices with hemorrhage multiplied by the slice thickness." (Kothari *et al.*, 1996)

Use of this formula often results in large volume estimation errors, particularly for large or irregularly shaped objects (Divani *et al.*, 2011; Wang *et al.*, 2009).

There is significant interest in semi-automatic computer models that can offer fast, low-bias, and precise lesion volume measurements (Kosior *et al.*, 2011). Among computer-based models, level set algorithms (Osher and Sethian, 1988b) have become a widely used method of determining lesion, organ, and bone volumes in recent years (Anandh *et al.*, 2014; Ivanovska *et al.*, 2016).

Semi-automatic segmentation algorithms are generally user 'seeded'. The reader places a seed on the image that triggers segmentation of an associated region of interest. Level set algorithms grow a seed placed within the region of interest based on pixel intensity to cover the entire lesion. Several parameters including the rate of growth and curvature of the segmented surface are controlled by local image properties and can also be adjusted by the reader (Whitaker, 1994). Level set algorithms are robust and flexible and prevent growth of the seed across weak, incidental connections into areas outside the lesion (Roberts *et al.*, 2010).

For example, Cates et al. (Cates *et al.*, 2004) demonstrated in a study of 9 meningioma or low grade glioma MRI images, that level set volume measurements produce results that are similar to those from hand tracings. Colliot et al. (Colliot *et al.*, 2006) found that measurement bias was low when they used level sets to measure cortical dysplasia lesions on MRI in 18 patients. Similarly, Saba et al. (Saba *et al.*, 2014) obtained a low bias using level sets to measure carotid artery wall thickness on MRI's of 10 patients; the level set volumes were compared to volumes calculated using manual tracings in both studies.

Level set algorithms require an enormous number of computations, consequently, being often too slow for practical clinical applications. To address this, we previously

developed a novel level set algorithm that leverages the massive parallelism of commodity graphical processing units (GPUs) (Roberts *et al.*, 2010). This algorithm is 14 times faster than the fastest previously reported parallel algorithm, and hundreds of times faster than serial algorithms on current central processing units. The algorithm had low bias and variability when segmenting tissues in a realistic brain phantom, an anatomically accurate three dimensional simulation of the human brain. Dang et al. (Dang *et al.*, 2013) reported that its speed, bias, and precision in measuring meningioma volume are superior to modified McDonald criteria and manual outlining on 25 contrast enhanced MR exams. However, the speed, bias, and precision of level set algorithms have not been tested using a large sample size of ICH and multiple clinician readers.

In this study, two vascular neurologists and two neuro-radiologists independently performed a test-retest experiment that involved repeated measurements of static, unchanging ICH volumes on CT. Measurements were made using both the level set algorithm and the ABC/2. We also estimated true ICH volumes using a consensus of repeated computer-assisted manual tracings (planimetrics) by one neurologist and one neuro-radiologist. Together, these data allowed us to estimate measurement bias, precision and acquisition time. To the best of our knowledge, this is the first time all three of these clinically relevant characteristics of ICH volume measurement have been estimated, analyzed and reported together. Bias and precision estimates presented here may aid to the design and execution of future clinical trials.

5.2 Methods

After IRB review, images for this study were obtained from an existing database of hemorrhagic stroke cases presenting to Mayo Clinic Hospital in Arizona between (date) and (date). Seventy-six cases with CT imaging available and noted intraparenchymal hematomas were selected. Exclusion criteria included extracerebral hemorrhage including intra-ventricular hematoma. Six cases were randomly withheld from the experimental dataset and used for training purposes.

Four operators, two neuro-radiologists and two vascular neurologists, collected level set and ABC/2 measurements in this study using custom developed GPU level set software [] and Osirix (Pixmeo, Geneva, Switzerland) respectively. The 70 non-training cases were each measured twice, by four operators, using both methods, for a total of 1120 measurements (70 cases x 4 operators x 2 methods x 2 repeats). One of the neurologists and one of the neuro-radiologists also manually outlined lesions using Osirix (70 cases x 2 operators x 2 repeats) in order to estimate true lesion volumes (described below).

In order to limit learning effects, repeated measurements of all cases in random order occurred after a minimum two-week delay. Manual tracings were done first followed by ABC/2 and finally level set. An additional level set correction process was introduced after initial data analysis. Figure 1a reviews the workflow for both the ABC/2 and level set measurement processes.

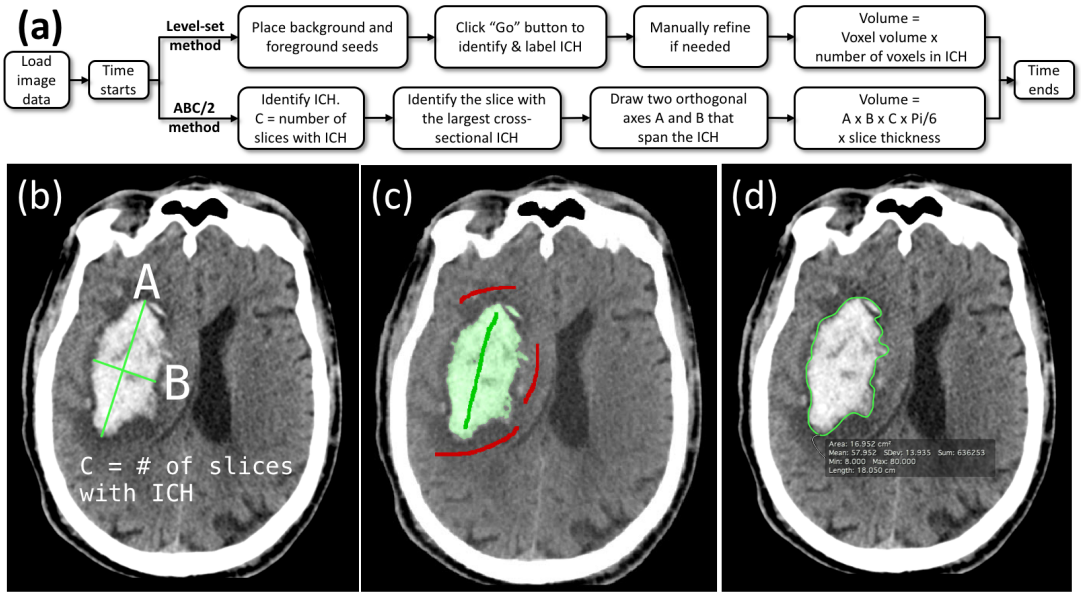

Figure 13. Hemorrhage measurement methods. A) a flowchart describing steps in the ABC/2 and level set measurement processes. b) Results from the ABC / 2 method; c) Results from the level set method. Green and red strokes were placed by the user. Green indicates desired tissue (foreground) and red indicates undesired tissue (background). The level set region grows in 3D after the user clicks "Go" (Figure 2); and, d) Manual outlining of the boundary in a single axial slice. This process was replicated in each axial slice with visible ICH. Each true volume was determined from 4 manual tracings – two operators each performed two repeated manual tracings. A minimum interval of 1 week between repeated tracings was required. Any voxel selected in at least 3 of the 4 manual tracings was labeled a "true" lesion voxel, and used to estimate the true lesion volume.

5.2.1 ABC/2 Method

The operator identified the axial slices with the largest lesion extent, then measured the two longest orthogonal diameters of the lesion using OsiriX ROI length tool (Fig 1b). The lesion volume was estimated as:

$$\text{Lesion volume} = \frac{A \times B \times C}{2},$$

where A and B are the length of the two diameters, and C is slice thickness multiplied by the number of slices where lesion was visible.

5.2.2 Level Set Method

The operator drew one or more green strokes in the object(s) to be segmented (ICH in our experiments), and one or more red strokes in surrounding tissue they did not want to include (surrounding brain parenchyma in our experiments). Strokes were placed on 2D axial slices (Fig 1c and Fig 2a). The operator then clicked a "Play" button to initiate object growth. Growth occurred in real-time and could be visualized in two or three dimensions interactively (Fig 2b-c). The algorithm naturally handles the simultaneous growth, merging, and separation of multiple distinctly seeded regions.

The surface propagation was guided by a K nearest neighbor (KNN) statistical model of the intensity values provided by the seed points, and a curvature parameter between 0 and 1 that determined the surface stiffness. In all cases the curvature parameter was initialized to 0.25. During, or following, region growth, the operator could adjust a slider to manipulate the curvature parameter. Higher values encouraged smoother, smaller volumes. Lower values encouraged larger, more irregular volumes. The effects of changes in the curvature parameter on the segmented volume occurred and were displayed in real-time.

The user had the ability to then review the segmentation in 2D, or in 3D by rotating the CT scan volume (Fig 2d). After the user accepted the results, the application calculated and recorded the ICH volume, operation time, and binary mask (segmentation) of the ICH.

A common issue in level set segmentation is 'leakage'. Artificial connections between blood and bone (both bright on CT) can be caused by partial volume effects between hemorrhage and skull (due in part to the large slice thickness of clinical image exams). This can be corrected either by increasing the level set surface stiffness

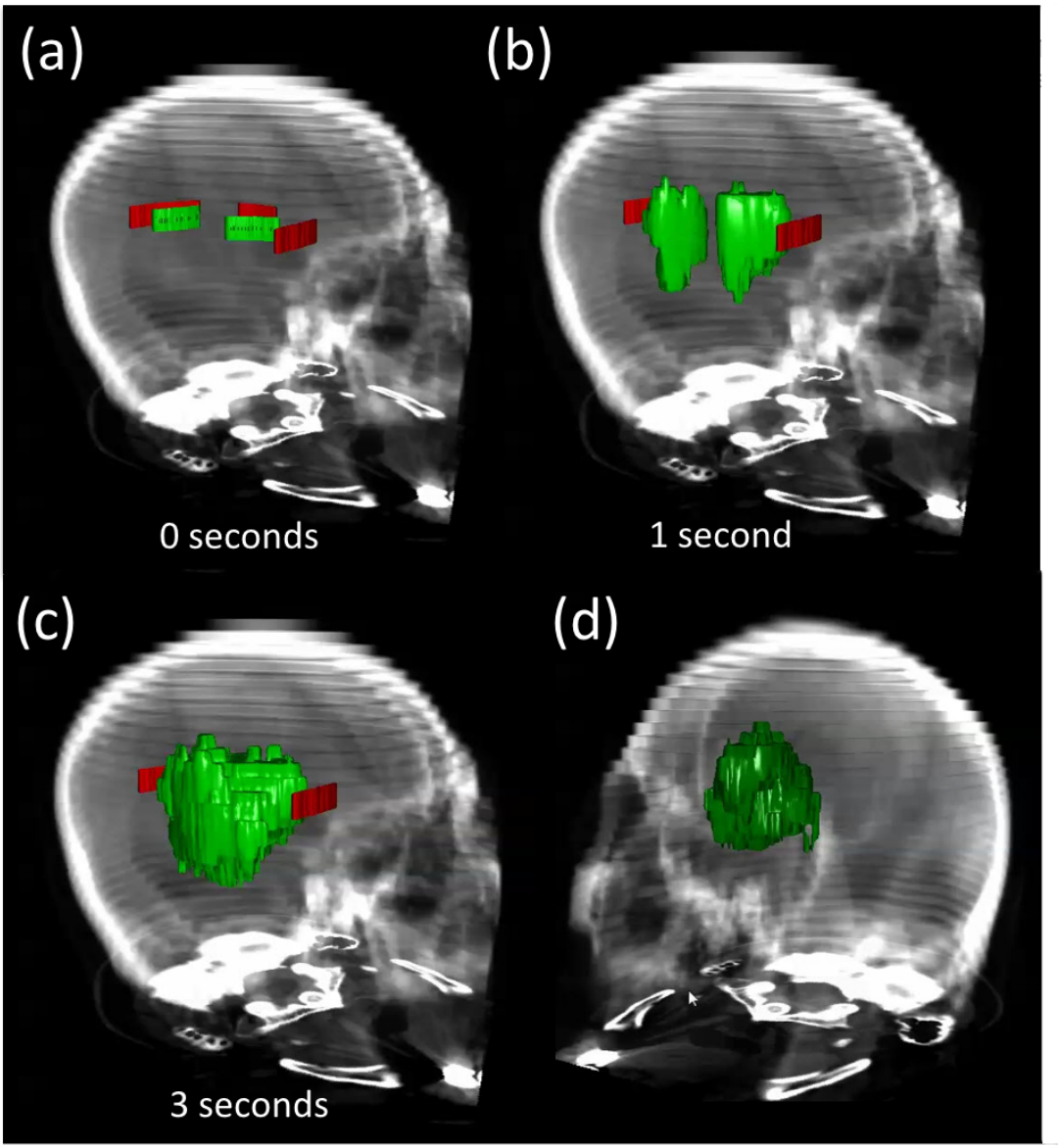

Figure 14. The level set ICH segmentation process in a CT scan with a 512 x 512 x 36 array. (a) The user placed foreground (green) and background (red) seeds in an axial slice (Figure 1c). Here the seeds are viewed in 3D. After the user clicks the "Go" button (not shown), the level set evolves to cover the ICH. In this example, a 41 ml ICH is segmented in 3 seconds (b-c). The final segmented region can be viewed in 2D, or in 3D by rotating the CT scan volume (d).

(causing the leak to retract back into the segmented region), or by removing the leaked regions with editing tools.

Accordingly, after all measurements were completed, the level set results were tested for outliers using the inter-quartile range outlier-labeling rule23. This test indicated that 24 / 560 measurements (4.3%) in 19 / 70 patients (27.1%) could be considered outliers (data not shown). It was determined that the default transparency level of the segmented region was such that operators found it difficult to detect leaks into bone (Figure 3a). Consequently, we asked all operators to perform a review of all of their level set measurements with the suggested workflow that they reduce the transparency of the green region until it was opaque and then review their measurements in 3D (Figure 3b). Readers were permitted to correct any segmentation either by editing or starting over. Both the initial and 'after review' level set measurements were recorded and compared to ABC/2 method.

5.2.3 True ICH Volume Estimation

The true ICH volume was estimated from manual tracings. One neurologist and one neuro-radiologist operator each independently manually traced the lesion on each axial slice in which it was apparent using the ROI pencil tool of OsiriX (Fig 1d). OsiriX was used to calculate the area in each slice and the lesion volume was calculated as:

$$\text{Lesion volume} = \text{slice thickness} \times \text{summation of lesion areas of all slices.}$$

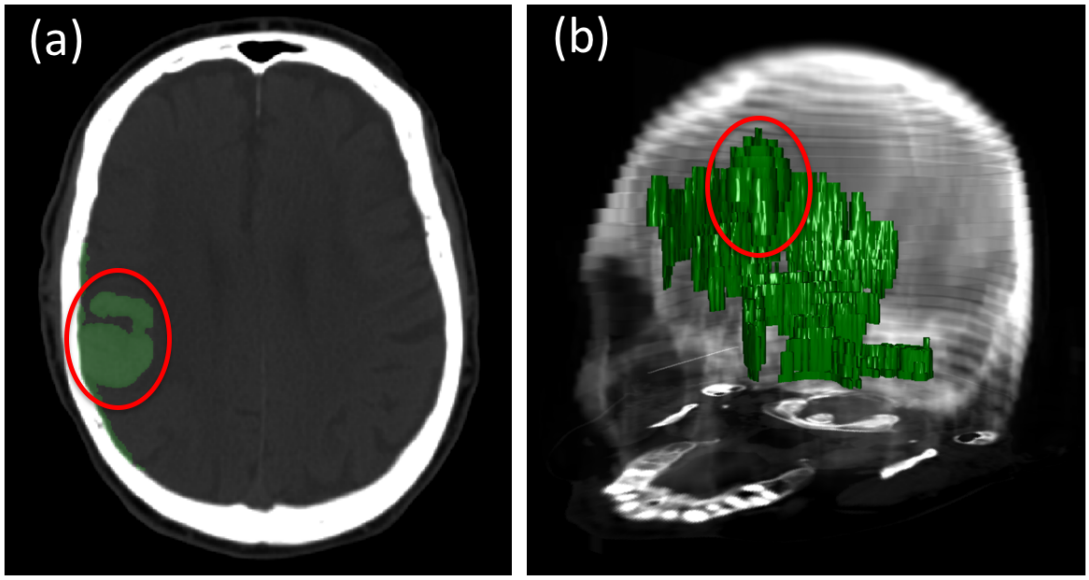

Figure 15. The level set algorithm may leak. An axial (a) and 3D view (b) of one segmented ICH lesion (enclosed in a red ellipse) along with level set leakage outside of the lesion. Leakage was caused by partial volume effects between hemorrhage and skull (due in part to the large slice thickness of the clinical imaging exams). This can be corrected either by increasing the level set surface stiffness, or by removing the leaked regions with editing tools. However, when the transparency of the segmented (green) region was high (as it was by default), operators found it difficult to detect leaks. Consequently we asked all operators to perform a blinded review of their level set segmentations. They performed this by interactively reducing the transparency of the green region until it was opaque, then viewing their segmentation in 3D (b). If, for any reason, they were unsatisfied with the labeled region, they were asked to re-segment the lesion. In the text we describe several enhancements to the tool that could be made to help operators detect leaks prior to saving the final segmented region.

Each true volume was estimated from four manual tracings by the two operators. Any voxel selected in at least three of the four manual tracings was labeled a "true" lesion voxel, and used to estimate the true lesion volume.

5.2.4 Statistical Analysis

We used methodology recommended by the Quantitative Imaging Biomarker Alliance to assess the uncertainty in volume measurement. In particular, we used a disaggregate approach to express uncertainty in terms of bias and precision. They define bias as "the difference between the average (expected value) of measurements made on the same object and its true value." They define precision as "the closeness of agreement between measured quantity values obtained by replicate measurements on the same or similar experimental units under specified conditions."

Precision is related to the variability in volume measurement. There are several ways it can be described numerically. Here, we chose to express precision as a clinically relevant threshold: the 95% minimum detectable change24 (MDC). Magnitude changes in volume $\leq$ to the MDC can be explained by variability in the measurement process alone, with 95% confidence. Therefore, magnitude changes in volume >MDC can be ascribed to lesion growth (or shrinkage) with a type II error rate of 5%.

We used a two-way random-effects analysis of variance (ANOVA) as previously described (Eliasziw *et al.*, 1994; Mitchell *et al.*, 1996), to estimate between- and within-operator variability for both measurement methods. These values were then used to calculate the between- and within-operator MDC for each method. The measurement bias of each method was determined by comparing the mean measured ICH volume to the estimated true ICH volume. The measurement time of the ABC/2 method was

determined by timing the readers as they measured the volumes. The measurement time of the level set method was recorded directly by the program.

5.3 Results

Summary statistics for our experiments are presented in the Table. The ABC/2 method produced larger volume estimates than the level set method. Level set measurements, after review, were smaller on average than the initial level set measurements. Operators modified 140 (of 560) level set measurements during the review process.

Table 8. Summary statistics for our experiments. All values are in mL, except N that is dimensionless. The values in the Level set (revisions) column indicate the change after operators reviewed their initial measurements. Negative values mean the after-review value was smaller than the initial value.

	ABC/2	Level set (initial)	Level set (after review)	Level set (revision)
N	560	560	560	140
Mean (std. dev.)	35.28 (56.24)	26.02 (40.35)	25.32 (39.76)	-2.2 (12.7)
1st Quartile	7.00	4.77	4.84	-1.43
Median	18.48	13.19	13.09	-0.01
3rd Quartile	36.82	28.87	27.47	0.40

Figure 4 shows the measurement bias (a), precision (b), and time (c) for the ABC/2 and level set (after review) methods. On average, the ABC/2 and level set methods produced measurements that were 45% and 3% larger than truth, respectively. (The slope of the line of best fit for ABC/2 and level set was 1.45 and 1.03, respectively.) The slope of the line of best fit for level set was not statistically different than 1.0, the line of perfect agreement between measured and true volumes ($p > 0.05$).

Between-operator MDC for ABC/2 was 13.94 mL. This was significantly higher (worse) than level set MDC before (11.93 mL, $p < 0.02$) and after (9.51 mL, $p < 0.02$) the operators reviewed for level set leakage.

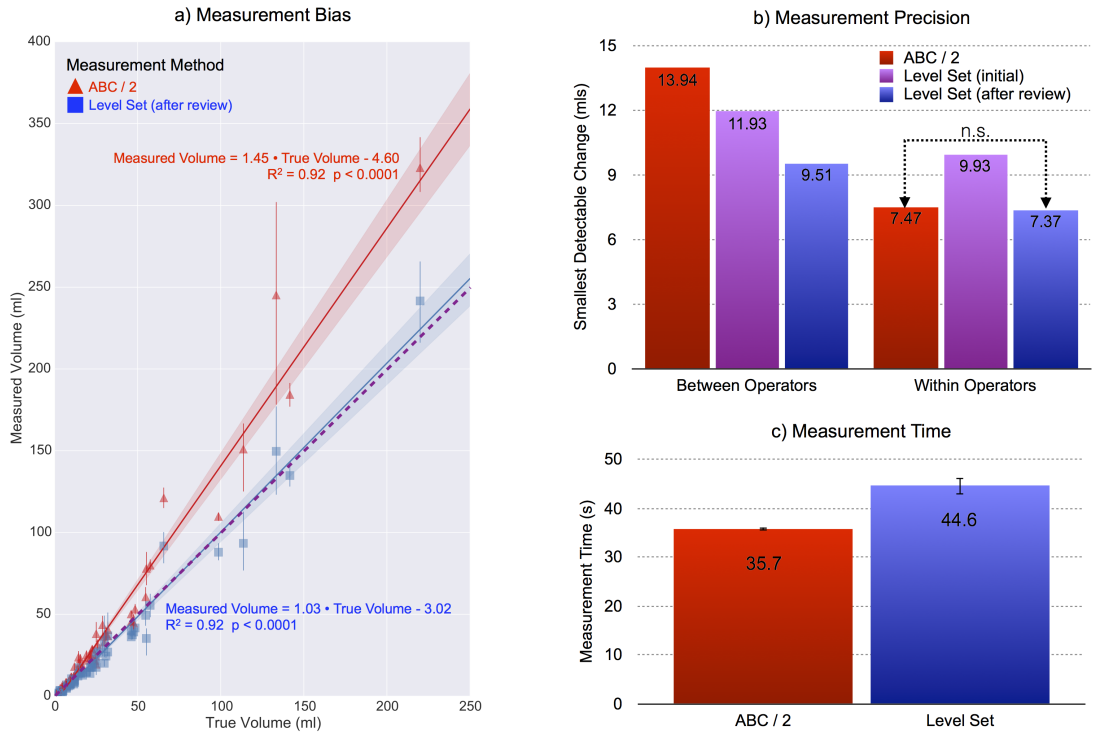

Figure 16. The: (a) bias; (b) precision; and, (c) time for ICH volume measurement. Measurements made using the ABC/2 method are shown in red. Those made using the level set are shown purple (initial) or blue (after review). (a) Bias: the dashed (purple) line indicates perfect agreement between measured and true volumes. Each true volume was determined from 4 manual tracings placed by experts (details in the text). The solid markers indicate the mean measured volume determined from 8 measurements (4 operators x 2 measurements / operator). The error bars indicate the 95% confidence interval for the mean measured volume. The solid lines show a linear regression fit through all measurements for each method. The shaded zone around each solid line indicates the 95% confidence interval for the slope of the line. Each line is labeled with its equation, the p-value from the linear regression, and the Coefficient of Determination describing goodness of fit (1.0 equals a perfect fit). (b) Precision: lower values are better. Changes in ICH volume less than the indicated value can be explained by measurement variability alone, with 95% confidence. The values in the "Between Operator" group are all statistically significantly different ($p < 0.02$). The "Level Set (initial)" value in the "Within Operators" group is significantly different than the other values ($p < 0.001$), which are not significantly different from each other ($p = 0.8$). (c) Time: each bar indicates the mean value from 560 measurements (70 ICH lesions x 4 operators x 2 repeated measurements). The error bars indicate the standard error of the mean. On average, the Level Set required an extra 8.9 seconds to measure ICH volume. This difference is statistically significantly different from 0 ($p < 0.0001$).

Within operator MDC for the ABC/2 method was 7.47 mL. This was significantly lower than the MDC for the initial group of level set measurements (9.93 mL, $p < 0.001$). However, after the operators reviewed and revised their level set measurements to eliminate leaks, level set MDC improved to 7.37 mL. This value was not statistically significantly different than the ABC/2 MDC ($p = 0.8$).

Measurement time is the total time required to complete the steps listed in Figure 1 (a), after the data was loaded. The average measurement times for the ABC/2 and level set (initial) methods were 35.7 s and 44.6 s, respectively. The average measurement time for repeating the level set measurements (after review) was not statistically significantly different than the average initial measurement time (data not shown). The level set method took an extra 8.9 seconds to complete, on average. This difference was significantly different from 0 ($p < 0.0001$).

5.4 Discussion

5.4.1 ICH Measurement Bias

To determine bias, true lesion volumes are required. These are often unknown in clinical experiments. A compromise is to estimate the true volume of each lesion using a consensus of expert segmentations in medical images. In our experiments, true ICH volume was estimated by a majority-voting scheme. A voxel was considered part of the true ICH if it was selected in at least 3 of 4 manual tracings performed by experts (two operators each performed two manual tracings of each lesion). Other schemes are also possible (Kessler *et al.*, 2015).

Several studies have reported that baseline ICH volume is highly predictive of

30-day morbidity and mortality (Broderick *et al.*, 1993; Hemphill *et al.*, 2001). It is also well acknowledged that the ABC/2 method tends to overestimate lesion volumes (Xu *et al.*, 2014; Webb *et al.*, 2015). In our study, the ABC/2 method overestimated ICH volume by 45%, on average. On the other hand, the mean level set measurements were no different than the estimated true ICH volumes. The low measurement bias of the level set algorithm could be important for future clinical trials that rely on baseline ICH volume for patient stratification.

5.4.2 ICH Measurement Precision

There was no significant difference in the within operator MDC between the ABC/2 and level set method (after review to check for leaks). However, the level set method had between operator MDC that was approximately 50% lower (better) than the ABC/2 method. This is important, since in many clinical scenarios it is difficult to ensure that a single individual performs all volume measurements.

A recent study by Dowlatshahi et al. (Dowlatshahi *et al.*, 2011) reported that hematoma expansion of 33% or 12.5 mL was highly predictive of poor outcome. The between-operator MDC of the ABC/2 method in our study was 13.94 mL. Consequently, an expansion of up to 13.94 mL measured this way can be explained by measurement variability alone (with 95% confidence) and may not correspond to true hematoma growth. The level set method, on the other hand, had a between-operator MDC of 9.51 mL. An expansion of 12.5 mL measured this way is too large to be explained by measurement variability alone. Consequently, it may be ascribed to ICH growth.

5.4.3 Comparisons to Previous Studies

A number of previous studies have compared the ABC/2 method to one or more computer-assisted methods for measuring ICH volume. For example, Divani et. al.32 compared ABC/2 to planimetric measurements made with MIPAV (NIH, Bethesda, MD) and to computer assisted segmentation using Analyze (Analyze Direct, Overland Park, KS) and Voxar 3D (Barco NV, Belgium). Their study focused on determining measurement bias. Measurements were made from a silicone phantom, and from simulated lesions implanted in cadaver brains. A variety of CT protocols were used. They found that ABC/2 produced large over estimations of lesion volume. They did not report measurement precision or time.

Freeman et al. (Freeman *et al.*, 2008) compared ABC/2 to Analyze in a series of 8 ICH patients treated with Warfarin. They used the Analyze measurements to estimate true lesion volumes. These were then used to estimate the measurement bias of ABC/2. They concluded that ABC/2 has small measurement bias for small lesions, but can produce large errors (both over and under estimation of true volume) for large or complex lesions. They did not exhaustively examine measurement time, but reported that the most accurate method of measuring lesion volume with Analyze required 20 – 30 minutes per study, after a steep learning curve. They did not report measurement precision.

Xu et. al. (Xu *et al.*, 2014) compared ABC/2 to computer-assisted measurements made using an intensity threshold algorithm in 3D Slicer (www.slicer.org) in 294 CT exams. Two physicians each made a single measurement of each lesion using each method. They used the Slicer measurements to estimate true lesion volumes. These were then used to estimate the measurement bias of ABC/2. They also divided the

lesions into groups based on size and shape, and compared the two measurement methods between groups. They concluded that ABC/2 over estimated lesion volumes, and that these errors increased with lesion volume, and with more complex shape. They reported that the average time required to measure lesion volume with 3D Slicer was 96 seconds. They did not report measurement precision or ABC/2 measurement times.

Kosior et al. (Kosior *et al.*, 2011) compared ABC/2 to computer-assisted measurements made using custom-developed software (Quantomo) that used an intensity threshold algorithm available as part of the Insight Segmentation and Registration Toolkit (ITK, NLM / NIH, Bethesda, MD). Kosior's study was focused on evaluating measurement precision. Like our study, they used a test-retest experiment and a two-way random-effects ANOVA to determine MDC for each method. They also had 4 operators each perform 2 repeated measures using each measurement method. They reported the within-operator MDCs for ABC/2 and Quantomo as 15.7 and 5.3 mL, respectively (vs. our values of 7.47 and 7.37 for ABC/2 and level set (after review), respectively). Their reported between-operator MDCs for ABC/2 and Quantomo were 37.0 and 8.1 mL, respectively (vs. our results of 13.94 and 9.51 mL, respectively).

There were several differences between their methods and ours. These include: 5 days between repeated measures (vs. 2 weeks in our study); and, measurement of 30 CT scans (vs. 70 in our study). It is also likely that their CT scan protocol was more consistent between patients than ours, since their patients were recruited from a clinical trial whereas ours were recruited from our general patient population over several years. Another potential difference: in Kosior's paper, all of the cross-sectional images of the brain do not have a visible skull, suggesting that a skull-stripping algorithm may have been used as a pre-processing step. The paper does not mention the presence

or absence of any pre-processing steps, however. CT scans in our study had no pre-processing performed. Skull-stripping, if applied, should reduce the variability of computer-assisted volume measurement by reducing the likelihood of leaks or growth from ICH into partial volume voxels that overlap blood and skull. Finally, they did not report measurement bias or times.

5.4.4 Level Set Leaks

One of the weaknesses of our study was the need to have operators review their initial level set measurements. This was performed in a blinded fashion – operators were not provided with any information on which studies may have had leaks. They were free to redo or edit any, all or none of their prior measurements. In total, 140 of the initial 560 segmentations were revised. The median result was to reduce the lesion volume by 0.01 mL. The mean effect was to reduce lesion volume by 2.2 mL. The impact on measurement precision was to reduce between-operator MDC from 11.93 to 9.51 mL, and to reduce within-operator MDC from 9.93 to 7.37 mL.

There are several modifications we could make to the measurement process to increase the visibility of level set leaks and/or reduce their likelihood. By default, the segmented region was highly translucent. This made level set growth into brightly saturated bone very difficult to see. Simply increasing the default opacity of the segmented region would address this limitation.

Next, we could place "high water" and "low water" indicators on the slider controlling surface stiffness. This would provide visual feedback to operators about the range over which they have adjusted surface stiffness during the current region growing process. This would help operators identify parameter values that allowed leakage to

occur. Increasing surface stiffness slightly causes the leaked region to quickly retract into the main body of the segmented region. The operator would then be able to lower stiffness to a point just above the "low water" mark on the slider to achieve an acceptable segmentation, without leaks.

We could also add an independent visual cue to indicate the size of the segmented (green) region, relative to the size of the image volume (or the size of all non-background voxels in the volume). Imagine, for example, a display showing a scaled down 3D representation of the segmented region embedded within a larger translucent cube (representing the image volume). The size of the embedded region could change, interactively, in proportion to the size of the segmented region. This would provide operators a visual indication of the size of the segmented region relative to the size of the image volume, without having to explicitly switch to a 3D view of the dataset or adjust the opacity of the segmented region.

Finally, we could also pre-process the image volumes to strip skulls. Several open-source algorithms are available to perform this task. This would cause delays in interpretation however.

5.5 Conclusion

Although an ICH volume measurement with the SegTool took 9 s longer to complete on average, compared with the traditional ABC/2 method, the SegTool was superior in all other clinically relevant aspects. The SegTool ICH volumetric measurements were not significantly different than the true ICH volumes, while ABC/2 overestimated ICH volume by 45%. The between-operator measurement variability with the SegTool was 50% less than with ABC/2. The SegTool was capable of detecting minimal clinically

important differences in ICH volume, whereas the ABC/2 was not. The SegTool appears to have measurement attributes superior to ABC/2 method and could be useful in clinical trials and clinical practice when intervention and prognostication rely upon accurate baseline ICH volume or upon detecting change in serial ICH volumetric measurements.

Chapter 6

DISCUSSION

This application obtains quantitative information from medical images. It extracts meaningful information and knowledge from unstructured medical imaging data and pave the way for further medical image analysis and decision making, which is the next step within biomedical imaging informatics. For instance, computer aided diagnosis (CAD) systems could use this information as input to help the physicians provide personalized patient care. Furthermore, researchers from other area of biomedical informatics could combine this quantitative image information with their own information to conduct research or make clinical decisions. For instance, along with genome data, this could be used to discover the correlation between radiology and genomics, potentially revolutionizes the bioinformatics field. In addition, clinical decision making systems could use this application and nature language process tools to better utilize the all kinds of data from electronic medical records (EMR). Last but not the least, when using the EMRs of a large population, it could be used to determine and discover screening methods so that abnormalities could be detected earlier.

6.1 Potential Impact on Clinical Practice

In Chapter 5, I compared this application with the standard ABC/2 method to measure ICH stroke volumes. The standard ABC/2 method is the most commonly used clinical method to estimate stroke volume, "where A is the greatest hemorrhage

diameter by CT, B is the diameter 90 degrees to A, and C is the approximate number of CT slices with hemorrhage multiplied by the slice thickness." Use of this equation often results in significant volume estimation errors, particularly for large or irregularly shaped objects. RECIST simplifies the ABC/2 method by measuring only one length (A) of up to 5 lesions from images, but applied to solid tumor size measurements.

6.1.1 Bias

True lesion volumes are required to determine bias. These are often unknown in clinical experiments. One compromise is to estimate the true volume of each lesion using a consensus of expert segmentations in medical images. In the experiments, true ICH volume was estimated by a majority-voting scheme. A voxel was considered part of the true ICH if it was selected in at least 3 of 4 manual tracings performed by experts (two operators each performed two manual tracings of each lesion). Other schemes are also possible.

Several studies have reported that baseline ICH volume is highly predictive of 30-day morbidity and mortality. It is also well acknowledged that ABC/2 tends to overestimate lesion volumes (Xu *et al.*, 2014; Webb *et al.*, 2015). In this study, ABC/2 overestimated ICH volume by 45% on average. In contrast, the mean of this application measurements were no different than the estimated true ICH volumes. The low measurement bias of this application could be important for future clinical trials that rely on baseline ICH volume for patient stratification.

6.1.2 Precision

There was no significant difference in the within-operator minimum detectable change (MDC) between ABC/2 and this application. However, this application had a between-operator MDC that was approximately 50% lower (better) than ABC/2. This is important because in many clinical scenarios it is difficult to ensure that a single individual performs all volume measurements.

A recent study by Dowlatshahi et al. (Dowlatshahi *et al.*, 2011) reported that hematoma expansion of 33%,or 12.5 mL, was highly predictive of poor outcome. The between-operator MDC of ABC/2 in the study was 13.94 mL. Consequently, an expansion of up to 13.94 mL measured this way can be explained by measurement variability alone (with 95% confidence) and may not correspond to true hematoma growth. This application, on the other hand, had a between-operator MDC of 9.51 mL. An expansion of 12.5 mL measured this way is too large to be explained by measurement variability alone. Consequently, it may be ascribed to ICH growth.

6.1.3 Acquisition Time

The average measurement times for ABC/2 and this application were 35.7 s and 44.6 s, respectively. This application required an extra 8.9 seconds on average to complete measurements. The 9 seconds delay could be explained that the operators use ABC/2 method much more regularly in their clinical practice, while only used this application to measure volumes after one to two weeks training. Nonetheless, it is still much faster than the manual delineation or other computer-assisted methods. Freeman et al. (Freeman *et al.*, 2008) compared ABC/2 to Analyze software in a

series of 8 ICH patients treated with Warfarin. They did not exhaustively examine measurement time, but reported that the most accurate method of measuring lesion volume with Analyze required 20 – 30 minutes per study, after a steep learning curve.

6.1.4 Future Work

6.1.4.1 Translation to iCORE Lab in Radiology

One of the places which could benefit from this application is the imaging Clinical Outcomes and Response Evaluation (iCORE) Lab at Mayo Clinic Arizona. iCORE provides comprehensive image management, quality control infrastructure, and rapid, reliable analysis for research trials. Standard RECIST measurements are performed by trained radiology technologists then reviewed by Mayo Clinic radiologists who have additional training in clinical trial measurement protocols. The measurements are made with Mint Lesion V3.0 (Mint Medical, Heidelberg Germany), a commercial, FDA-cleared software designed for this purpose. These provide a great environment to validate this application. To integrate this application into the core lab workflow, more models need to be implemented. First, we need to establish the connection to image database in the radiology picture archiving and communication system (PACS). Then after the measurement, the volume and/or the segmentation (saved as a binary mask) should be send back to PACS. In addition, a reporting model to present the measurement numerically and graphically is helpful to clinicians and patients.

6.1.4.2 Clinical Research Studies

We conducted one clinical research study on ICH stroke with 76 patients. It is a comprehensive and time consuming process, but more retrospective validation studies should be conducted to evaluate the bias, precision, and acquisition time of this application and to advance quantitative imaging technology. New studies should follow the test-retest framework described in Chapter 5. For instance, a good validate study could be applied to glioblastoma (GBM) tumor volume measurement. Volumetric and parameters of GBM tumor have been used to predict the prognosis and to determine treatment failure (Iliadis *et al.*, 2012; Grabowski *et al.*, 2014). Contrast-enhancing residual tumor volume (CERTV) alone significantly predicts survival after GBM resection, reflecting the pathobiology of GBM. However, this quantitative information is rarely recorded in clinical practice. Additional validation study is needed to prove the concept of adopting this application to such case

6.2 Potential Impact on Biomedical Research

In addition to volume information, this application obtains the entire morphological information of the anatomical structures, such as the shape and location. These are valuable to other biomedical image based research.

6.2.1 Extract GBM Brain Tumor

The Mathematical Neuro-Oncology (MNO) lab at Mayo Clinic focuses on cancer research to build mathematical models to predict cancer patient's prognosis and to

optimize treatment. One system in their lab monitors the longitudinal patient's tumor size from multiple MRI data along with the treatments. The goal is to provide personalized precision treatment choices to improve the lives of patients with brain cancer (Neal *et al.*, 2013b,a). As shown from the figure below, a patient had about 50 measurements during 3 years of cancer treatment. MNO uses a measurement team to delineate tumor boundaries from T1, T2, and FLAIR MRI data. Currently the operators of the measurement team manually trace tumors. This application can significantly increase their productivity. More data could lead to more research discoveries and better outcome of cancer patients.

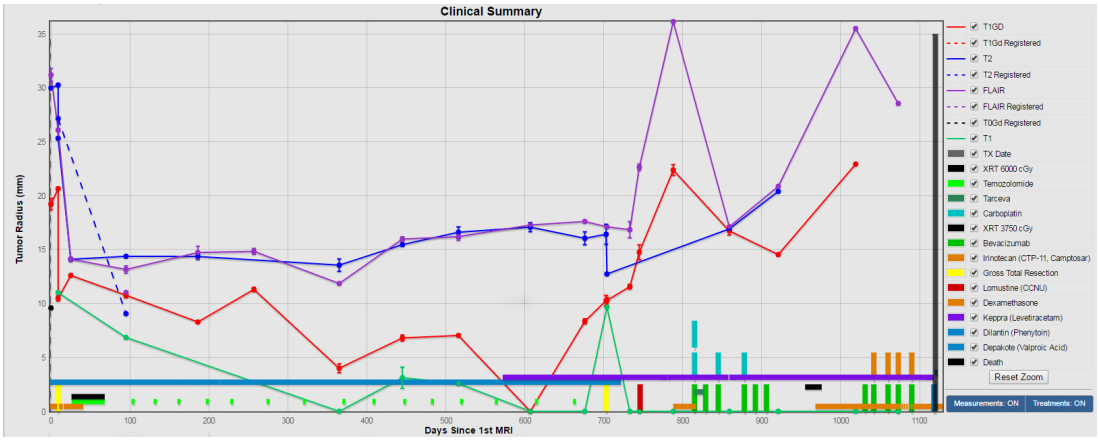

Figure 17. Brain tumor patient graph with tumor radius and treatments.

We conducted a preliminary study to evaluate this application's performance on segmenting brain tumors. This application achieved comparable results as the state-of-the-art methods when segmenting whole tumor using BRATS 2013 MRI data, and achieved acceptable results comparing to human labeled references using MNO's own MRI data.

6.2.2 Personalized Neurosurgery

Another research area is neurosurgery, which requires the accurate structure of the target region. This application had been frequently used by a team of neurosurgery researchers at Mayo Clinic Arizona to extract vessels and aneurysms from brain CT angiography. First, the surgeon would prefer a 3D visualization of the aneurysm and surrounding vessels to plan the surgery. Then the binary segmentations were converted to 3D meshes. With the current development of 3D printing, the 3D meshes were used to 3D print the vessels and aneurysm. 3D printing technology brings anatomic structures showing on the digital screen to the palms of the surgeons. It allows the surgeons to examine and to simulate and practice prior to the real surgery. For aneurysm, they can test multiple treatment plans and identify the best way to treat the specific patient. This can shorten the real surgery time and improve the healthcare quality. In addition to surgery planning, these 3D printed anatomical structures provide great education to both patients and medical students/residents.

For a preliminary study with one CT angiography from a patient, we compared the vessel and aneurysm segmentation using this method to a human labeled reference and found the largest difference between the two surfaces was less than 1 mm, which is subvoxel. But, the human labeling process took hours to only obtain a small region while this application could extract much more vessels within minutes.

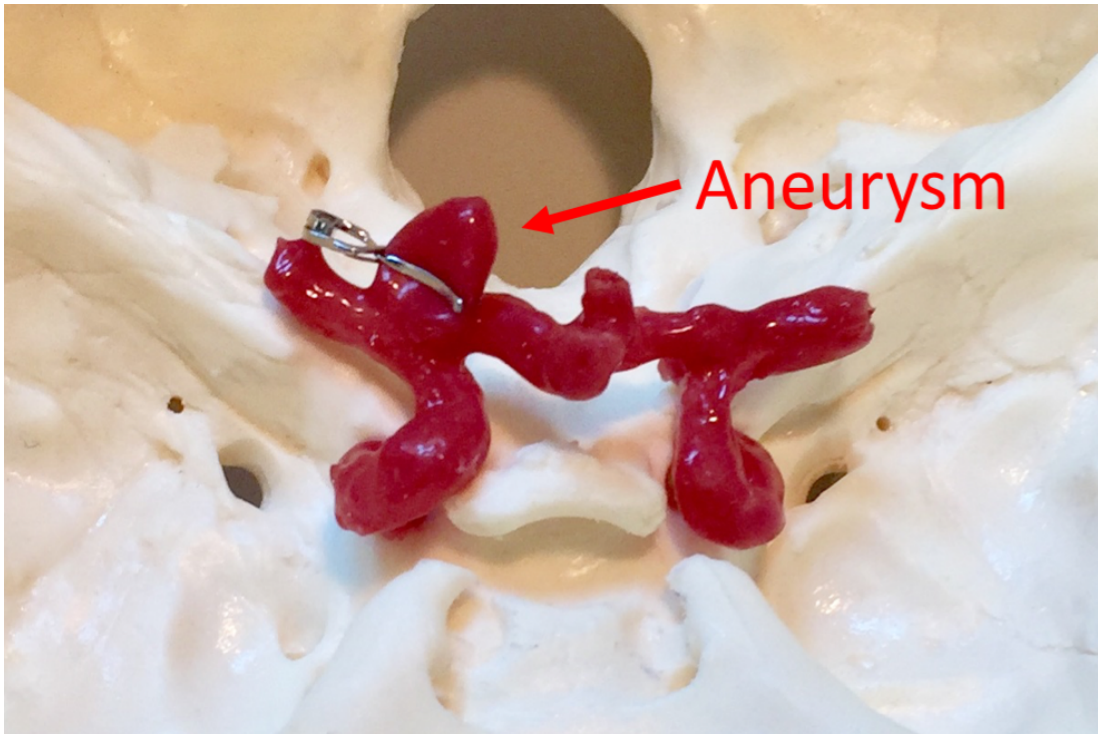

Figure 18. A photo of the 3D printed vessel and aneurysm using silicon.

6.2.3 Future Work

6.2.3.1 Automatic Seed Generating

One limitation of this application is that the users need to place foreground and background seed points. After the validation study showing this application is able to obtain accurate and precise results, auto seed generating becomes the next step in order to eliminate the human inputs and fully automate the entire process. This can reduce the variability associated with different operators (researchers or clinicians). One potential way to achieve auto seeding is to identify statistical image feature model and using the voxels with high probabilities belonging to the background and foreground as background and foreground seed points, respectively. In addition, registering atlas

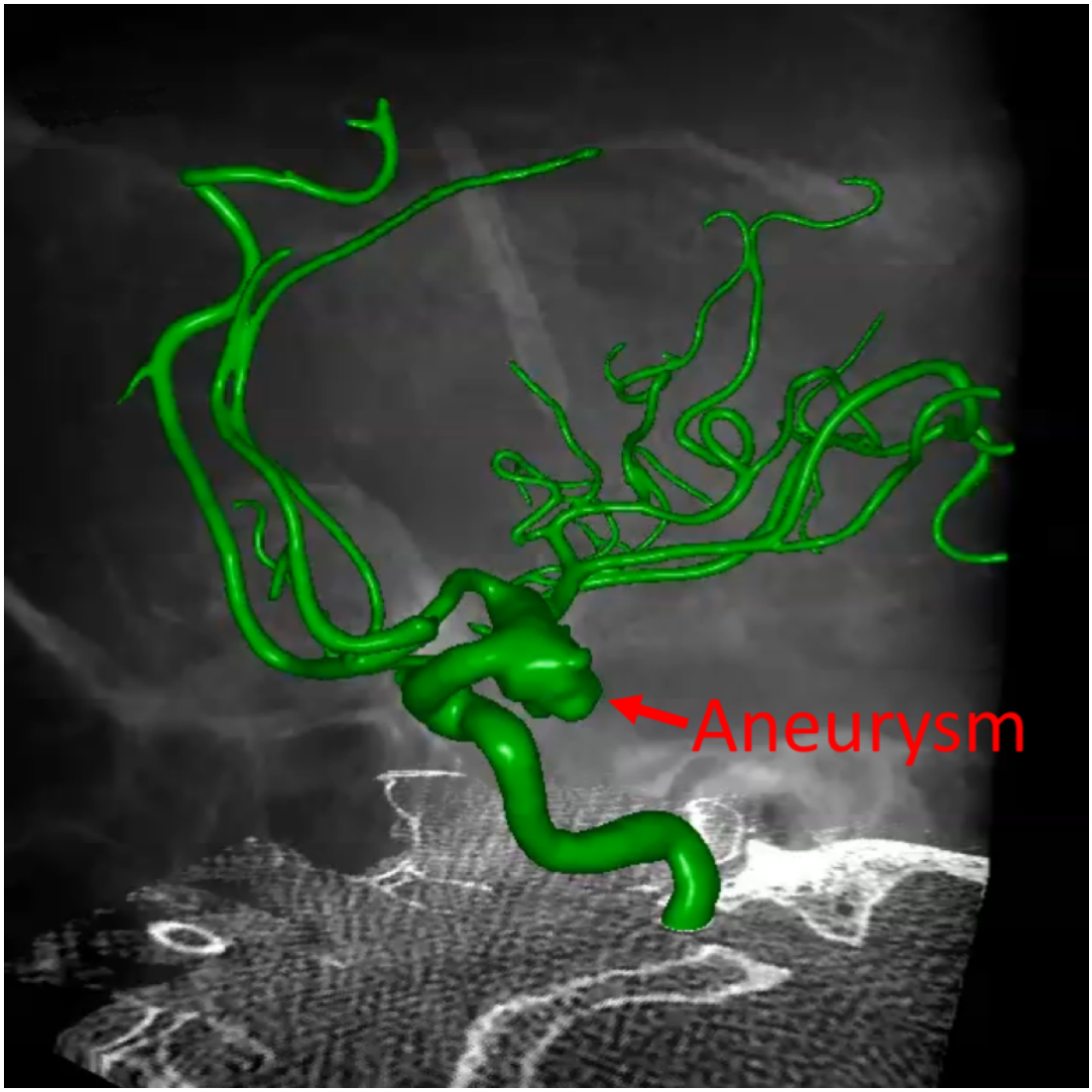

Figure 19. A 3D visualization of the segmentation of neurovascular and anuerysm from CT angiography using this application.

could be used to create background seed points if we are trying to segment abnormality from a healthy anatomy, such as registering brain atlas to a patient with brain tumor to find healthy tissues and sampling these voxels as background seed points to extract brain tumor. Then the user only needs to place foreground seed points to mark the tumor region.

6.2.3.2 Using Different Framework

We showed how to segment MS lesions in previous chapter (4). We proposed a new iterative framework, which integrates the stand alone application,to obtain a complicated task. We segmented MS lesions with multiple steps by segmenting different regions using different combination of seeds from multiple contrast MRI images. It is very difficult to achieve that with one pass of segmentation. This study shows the potential of this application being used in different ways to solve complex problems and encourages future researches to extend this application to other frameworks.

6.2.3.3 Embedding to Open Source Softwares

Translation to an open source research softwares, like 3D slicer (Fedorov *et al.*, 2012). First is separating the core algorithms from this application. The core includes two parts: the speed computation and GPU level set method. The speed function requires 1) one or more (coregistered) medical images and 2) user labeled seed points as inputs. All of these inputs could be handled by Slicer. The speed function generates a speed volume as output which could be stored as an ITK float image. On the other

hand, the GPU level set part takes 1) a speed volume, 2) foreground labels (where the growth starts), and 3) a tuning parameter (weight of surface smoothness) as inputs. The tuning parameter could be set by users with a Qt slider in Slicer. The GPU level set generates a binary mask as output, which could be stored as an ITK UInt8 image.

The second step is linking it to Slicer. Slicer provides users the abilities of loading and viewing medical images, labeling seeds, tuning parameter, and saving feature volume and binary mask with its IO, user interaction, and visualization modules.

After the integration, we hope the extension could achieve the following three scenarios:

1. Using speed function only: generating speed volume based on user labeled seeds and images as inputs. The output ITK image could be used for data visualization or colormap, in addition to guide GPU level set growth.
2. Using GPU level set only: generating a binary mask with a user specified parameter, foreground seeds (could load from a file), and speed volume (could be loaded from a file) as inputs. This is for general image segmentation task. And users could load their own precomputed speed volume.
3. Using speed function guided GPU level set: generating a binary mask with images, user labeled seeds, and user tuned parameter as inputs. This is the process I demoed during my presentation. Note that all the scenarios could be executed in console by passing parameter and file names of images, seed files, and speed file as arguments.

Chapter 7

CONCLUSION

Precise and unbiased measurement of tumor/lesion volume is important for cancer staging and therapeutic assessment. The current methods used in clinical practice emphasize on efficiency, but suffer from low bias and precision. In the other hand, computational algorithms focus on improving the accuracy, but are often time consuming and cumbersome to use. In addition, most of them lack validation studies to evaluate their practical impact to healthcare. All of these hinder the translation of advanced methods from benchside to bedside, thus not able to increase the healthcare outcome.

In this dissertation, I propose a user interactive image application to rapidly extract accurate quantitative information of abnormalities (tumor/lesion) from multi-spectral medical images, such as measuring brain tumor volume from MRI. This user interactive image segmentation application has a simple and intuitive graphical user interface. It requires the users to place foreground and background seed points, where foreground seeds label the desired region and background seeds label the undesired region. Theses seeds are used to generate prior image features to guide the surface propagation. Then the users can simply start the method to extract the targeted region. Only one parameter needs to be specified in order to adjust the surface smoothness. It could be empirically initialized based on image noise level to obtain optimal results and can be tuned by the users in real time. In addition, a comprehensive workflow is presented to validate image quantitative methods for clinical studies.

This application has been evaluated and validated in multiple cases, including

segmenting healthy brain white matter from MRI and brain lesions from CT or MRI. A retrospective validation study is conducted on measuring intracerebral hemorrhage volume from CT scan following this workflow. Such validations have not been done for most of the advanced computational methods.

These evaluations show that this application is capable to achieve low bias and high precision on volume measurement of ROIs. In particular, the clinical validation study with a large dataset demonstrates that this application appears to have measurement attributes superior to the current practice method in terms of accuracy and reliability without a significant delay in acquisition time. In other words, it could be useful to the clinical trials and clinical practice, especially when intervention and prognostication rely upon accurate baseline lesion volume or upon detecting change in serial lesion volumetric measurements. More clinical proof-of-concept studies should be conducted in order to validate the performance and to advance the role of quantitative image metrics in patient care.

Additionally, other biomedical researches, which require extracting information from medical images, could greatly benefit from this application also. As a general image segmentation tool, this application provides structural and morphological information in addition to the quantitative volumetric information. This could be used in researches such as predicting tumor growth, surgery planning. In order to reach a broader research community, this application could be embedded into popular open-source medical image platforms.

REFERENCES

Anandh, K. R., C. M. Sujatha and S. Ramakrishnan, "Atrophy analysis of corpus callosum in Alzheimer brain MR images using anisotropic diffusion filtering and level sets", in "2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014", pp. 1945–1948 (2014).

Berger, A., "How does it work?: Magnetic resonance imaging", BMJ: British Medical Journal **324**, 7328, 35 (2002).

Boykov, Y. and M.-P. Jolly, "Interactive graph cuts for optimal boundary amp; region segmentation of objects in n-d images", in "Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on", vol. 1, pp. 105 –112 vol.1 (2001).

Boykov, Y. and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision", Pattern Analysis and Machine Intelligence, IEEE Transactions on **26**, 9, 1124 –1137 (2004).

Brinkley, J. F. and C. Rosse, "Imaging and the Human Brain Project: a review.", Methods of information in medicine **41**, 245–260 (2002).

Broderick, J., T. Brott, J. Duldner, T. Tomsick and G. Huster, "Volume of Intracerebral Hemorrhage A Powerful and Easy-to-Use Predictor of 30-Day Mortality", Stroke. **24**, 7, 987–93 (1993).

Brosch, T., Y. Yoo, L. Y. Tang, D. K. Li, A. Traboulsee and R. Tam, "Deep convolutional encoder networks for multiple sclerosis lesion segmentation", in "MICCAI", pp. 3–11 (Springer, 2015).

Buzug, T. M., *Computed tomography: from photon statistics to modern cone-beam CT* (Springer Science & Business Media, 2008).

Caselles, V., F. Catt, T. Coll and F. Dibos, "A geometric model for active contours in image processing", Numerische Mathematik **66**, 1–31, URL http://dx.doi.org/10.1007/BF01385685 (1993).

Cates, J. E., A. E. Lefohn and R. T. Whitaker, "GIST: an interactive, GPU-based level set segmentation tool for 3D medical images.", Medical Image Analysis **8**, 217–231 (2004).

Cocosco, C. A., V. Kollokian, R. K.-S. Kwan, G. B. Pike and A. C. Evans, "Brainweb: Online interface to a 3d mri simulated brain database", NeuroImage **5**, 425 (1997).

Collins, D. L., A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes and A. C. Evans, "Design and construction of a realistic digital brain phantom.", IEEE transactions on medical imaging **17**, 3, 463–8, URL http://www.ncbi.nlm.nih.gov/pubmed/9735909 (1998).

Colliot, O., T. Mansi, N. Bernasconi, V. Naessens, D. Klironomos and A. Bernasconi, "Segmentation of focal cortical dysplasia lesions on MRI using level set evolution", NeuroImage **32**, 4, 1621–1630 (2006).

Cootes, T. F., G. J. Edwards and C. J. Taylor, "Active appearance models", in "IEEE Transactions on Pattern Analysis and Machine Intelligence", pp. 484–498 (Springer, 1998).

Cootes, T. F., C. J. Taylor, D. H. Cooper and J. Graham, "Active shape models their training and application", Comput. Vis. Image Underst. **61**, 1, 38–59, URL http://dx.doi.org/10.1006/cviu.1995.1004 (1995).

Dang, M., J. Modi, M. Roberts, C. Chan and J. R. Mitchell, "Validation study of a fast, accurate, and precise brain tumor volume measurement.", Computer methods and programs in biomedicine **111**, 2, 480–7 (2013).

Davis, S. M., J. Broderick, M. Hennerici, N. C. Brun, M. N. Diringer, S. A. Mayer, K. Begtrup and T. Steiner, "Hematoma growth is a determinant of mortality and poor outcome after intracerebral hemorrhage", Neurology **66**, 8, 1175–1181 (2006).

Demirkaya, O., M. H. Asyali and P. K. Sahoo, *Image Processing with Matlab: Applications in Medicine and Biology* (CRC Press Taylor & Francis Group Boca Raton, 2009).

Divani, A. A., S. Majidi, X. Luo, F. G. Souslian, J. Zhang, A. Abosch and R. P. Tummala, "The ABCs of accurate volumetric measurement of cerebral hematoma", Stroke **42**, 6, 1569–1574 (2011).

Dowlatshahi, D., A. M. Demchuk, M. L. Flaherty, M. Ali, P. L. Lyden and E. E. Smith, "Defining hematoma expansion in intracerebral hemorrhage: Relationship with patient outcomes", Neurology **76**, 14, 1238–1244 (2011).

Duda, R. O., P. E. Hart and D. G. Stork, *Pattern Classification (2Nd Edition)* (Wiley-Interscience, 2000).

Eklund, A., P. Dufort, D. Forsberg and S. M. Laconte, "Medical image processing on the GPU - Past, present and future.", Medical image analysis **17**, 8, 1073–1094, URL http://www.ncbi.nlm.nih.gov/pubmed/23906631 (2013).

Eliasziw, M., S. L. Young, M. G. Woodbury and K. Fryday-Field, "Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example.", Physical therapy **74**, 777–788 (1994).

Falcao, A. X., J. K. Udupa, S. Samarasekera, S. Sharma, B. E. Hirsch and R. de A. Lotufo, "User-steered image segmentation paradigms: Live wire and live lane", Graphical Models and Image Processing **60**, 4, 233 – 260, URL http://www.sciencedirect.com/science/article/pii/S1077316998904750 (1998).

Fedorov, A., R. Beichel, J. Kalpathy-Cramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper and R. Kikinis, "3D Slicer as an image computing platform for the Quantitative Imaging Network", Magnetic Resonance Imaging **30**, 9, 1323–1341 (2012).

Feigin, V. L., C. M. Lawes, D. A. Bennett, C. S. Anderson, C. Sarti, D. Rastenyte, Z. Cepaitis, J. Tuomilehto, C. Murray, A. Lopez, WHO, WHO, M. Foulkes, P. Wolf, T. Price, J. Mohr, D. Hier, J. Caro, K. Huybrechts, I. Duchesne, WHO, R. Bonita, A. Stewart, R. Beaglehole, P. Thorvaldsen, K. Kuulasmaa, A. Rajakangas, D. Rastenyte, C. Sarti, L. Wilhelmsen, R. Bonita, R. Bonita, R. Beaglehole, K. Asplund, C. Sudlow, C. Warlow, C. Warlow, K. Khaw, J. Posner, K. Gorman, A. Woldow, J. Kurtzke, J. Mas, M. Zuber, A. Viriyavejakul, S. Hatano, C. Sudlow, C. Warlow, O. Ahmad, C. Boschi-Pinto, C. Murray, R. Lozano, M. Inoue, A. s. o. r. a. n. W. world Standard, A. Thrift, H. Dewey, R. Macdonell, J. McNeil, G. Donnan, K. Jamrozik, R. Broadhurst, N. Lai, G. Hankey, P. Burvill, C. Anderson, H. Jorgensen, A. Plesner, P. Hubbe, K. Larsen, J. Stewart, R. Dundas, R. Howard, A. Rudd, C. Wolfe, H. Numminen, M. Kotila, O. Waltimo, K. Aho, M. Kaste, D. Smadja, P. Cabre, F. May, et Al., Y. Morikawa, H. Nakagawa, Y. Naruse, et Al., P. Kolominsky-Rabas, M. Weber, O. Gefeller, B. Neundoerfer, P. Heuschmann, P. Kolominsky-Rabas, C. Sarti, P. Heuschmann, C. Graf, S. Siemonsen, B. Neundoerfer, et Al., K. Vemmos, M. Bots, P. Tsibouris, et Al., G. Lauria, M. Gentile, G. Fassetta, et Al., A. Carolei, C. Marini, M. D. Napoli, et Al., R. Bonita, J. Broad, R. Beaglehole, H. Ellekjaer, J. Holmen, B. Indredavik, A. Terent, V. Feigin, D. Wiebers, Y. Nikitin, W. O'Fallon, J. Whisnant, L. Mihalka, V. Smolanka, B. Bulecza, S. Mulesa, D. Bereczki, R. Bonita, N. Solomon, J. Broad, M. Bots, S. Looman, P. Koudstaal, A. Hofman, A. Hoes, D. Grobbee, A. Nicoletti, V. Sofia, S. Giuffrida, et Al., M. Mittelmark, B. Psaty, P. Rautaharju, et Al., J. Geddes, J. Fear, A. Tennant, A. Pickering, M. Hillman, M. Chamberlain, P. O'Mahony, R. Thomson, R. Dobson, H. Rodgers, O. James, S. Lindeberg, B. Lundh, Z. Huang, T. Chiang, T. Lee, M. Prencipe, C. Ferretti, A. Casini, M. Santini, F. Giubilei, F. Culasso, F. Bermejo, S. Vega, J. Morales, et Al., P. Thorvaldsen, M. Davidsen, H. Bronnum-Hansen, M. Schroll, R. Brown, J. Whisnant, J. Sicks, W. O'Fallon, D. Wiebers, V. Feigin, D. Wiebers, J. Whisnant, W. O'Fallon, B. Stegmayr, K. Asplund, P. Wester, A. Czlonkowska, D. Ryglewicz,

T. Weissbein, M. Baranska-Gieruszczak, D. Hier, M. Rodrigues, M. Noronha, M. Vieira-Dias, et Al., R. Walker, D. McLarty, G. Masuki, et Al., T. Banerjee, C. Mukherjee, A. Sarkhel, S. Das, K. Sanyal, P. Thorvaldsen, K. Asplund, K. Kuulasmaa, A.-M. Rajakangas, M. Schroll, T. Ingall, K. Asplund, M. Mahonen, R. Bonita, G. Petty, R. Brown, J. Whisnant, J. Sicks, W. O'Fallon, D. Wiebers, K. Aho, A. Reunanen, A. Aromaa, P. Knekt, J. Maatela, N. Bharucha, E. Bharucha, A. Bharucha, A. Bhise, B. Schoenberg, H. Hu, F. Chu, B. Chiang, et Al., S. Li, B. Schoenberg, C. Wang, X. Cheng, C. Bolis, K. Wang, C. Paschalis, P. Polychronopoulos, N. Makris, X. Kondakis, T. Papapetropoulos, S. Razdan, R. Koul, A. Motta, S. Kaul, S. Sorensen, G. Boysen, G. Jensen, P. Schnohr, K. Urakami, M. Igo, K. Takahashi, T. Wyller, E. Bautz-Holter, J. Holmen, A. Terént, F. Linn, G. Rinkel, A. Algra, J. van Gijn, J. Tuomilehto, C. Sarti, E. Narva, et Al., J. Okwumabua, B. Martin, J. Clayton-Davis, C. Pearson, R. Sacco, B. Boden-Albala, R. Gan, et Al., R. Bonita, J. Broad, R. Beaglehole, T. Truelsen, R. Bonita, K. Jamrozik, R. Bonita, R. Beaglehole, K. Tilling, J. Sterne, C. Wolfe, K. Asplund, R. Bonita, K. Kuulasmaa, et Al., D. Schottenfeld, M. Eaton, S. Sommers, D. Alonso, C. Wilkinson, H. Cameron, E. McGoogan, M. Britton, J. Bamford, P. Sandercock, M. Dennis, et Al., M. Alter, E. Sobel, R. McCoy, et Al., C. Sarti, J. Tuomilehto, J. Sivenius, et Al., B. Stegmayr, K. Asplund, R. Malmgren, C. Warlow, J. Bamford, P. Sandercock, R. Bonita, J. Broad, N. Anderson, R. Beaglehole, T. Thom, K. Jamrozik, J. Reitsma, M. Limburg, J. Kleijnen, G. Bonsel, J. Tijssen, R. Gillum, C. Sempos, O. Ahmed, T. Orchard, R. Sharma, H. Mitchell, E. Talbot, J. Fang, M. Alderman, J. Tu, Y. Hong, M. Bots, X. Pan, A. Hofman, D. Grobbee, H. Chen, G. Howard, T. Craven, L. Sanders, G. Evans, P. Heidenreich, M. McClellan, M. Rosen, L. Alfredsson, N. Hammar, T. Kahan, C. Spetz, A. Ysberg, I. Bata, R. Gregor, B. Eastwood, H. Wolf, F. Spencer, T. Meyer, R. Goldberg, et Al., M. Kirchhoff, M. Davidsen, H. Bronnum-Hansen, et Al., R. Goldberg, J. Yarzebski, D. Lessard, J. Gore, J. Volmink, J. Newton, N. Hicks, et Al., W. Rosamond, L. Chambless, A. Folsom, et Al., J. Brophy, P. Immonen-Raiha, M. Arstila, J. Tuomilehto, et Al. and R. Beaglehole, "Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century", The Lancet Neurology **2**, 1, 43–53 (2000).

Fitzpatrick, J. M. and M. Sonka, *"Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis (SPIE Press Monograph Vol. PM80)"* (SPIE–The International Society for Optical Engineering, 2000), 1 edn.

Flaherty, M. L., M. Haverbusch, P. Sekar, B. Kissela, D. Kleindorfer, C. J. Moomaw, L. Sauerbeck, A. Schneider, J. P. Broderick and D. Woo, "Long-term mortality after intracerebral hemorrhage", Neurology **66**, 8, 1182–1186 (2006).

Freeman, W. D., K. M. Barrett, J. M. Bestic, J. F. Meschia, D. F. Broderick and T. G. Brott, "Computer-assisted volumetric analysis compared with ABC/2 method

for assessing warfarin-related intracranial hemorrhage volumes", Neurocritical Care **9**, 3, 307–312 (2008).

Freire, P. G. L. and R. J. Ferrari, "Automatic iterative segmentation of multiple sclerosis lesions using Student's t mixture models and probabilistic anatomical atlases in FLAIR images", Computers in Biology and Medicine **73**, 10–23 (2016).

García-Lorenzo, D., S. Francis, S. Narayanan, D. L. Arnold and D. L. Collins, "Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging", Medical Image Analysis **17**, 1, 1–18 (2013).

García-Lorenzo, D., J. Lecoeur, D. L. Arnold, D. L. Collins and C. Barillot, "Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts.", MICCAI **12**, Pt 2, 584–91 (2009).

Ge, Y., R. I. Grossman, J. S. Babb, M. L. Rabin, L. J. Mannon and D. L. Kolson, "Age-Related Total Gray Matter and White Matter Changes in Normal Adult Brain. Part I: Volumetric MR Imaging Analysis", AJNR Am. J. Neuroradiol. **23**, 8, 1327–1333, URL http://www.ajnr.org/content/23/8/1327.full (2002).

Geremia, E., O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi and N. Ayache, "Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images.", NeuroImage **57**, 2, 378–90 (2011).

Giedd, J. N., J. Blumenthal, N. O. Jeffries, F. X. Castellanos, H. Liu, A. Zijdenbos, T. Paus, A. C. Evans and J. L. Rapoport, "Brain development during childhood and adolescence: a longitudinal MRI study.", Nature neuroscience **2**, 10, 861–3, URL http://dx.doi.org/10.1038/13158 (1999).

Goldstein, L. B., "CLINICIAN ' S CORNER Is This Patient Having a Stroke ? EXAMINATION OF PATIENTS WITH SUSPECTED STROKE", Stroke **293**, 19, 2391–2402, URL http://stroke.ahajournals.org/cgi/content/abstract/37/9/2378 (2005).

Grabowski, M. M., P. F. Recinos, A. S. Nowacki, J. L. Schroeder, L. Angelov, G. H. Barnett and M. a. Vogelbaum, "Residual tumor volume versus extent of resection: predictors of survival after surgery for glioblastoma", Journal of Neurosurgery **v**, November, 1–9 (2014).

Grady, L., "Random walks for image segmentation", IEEE Trans. on Pattern Analysis and Machine Intelligence **28**, 11, 1768–1783 (2006).

Hemphill, J. C., D. C. Bonovich, L. Besmertis, G. T. Manley and S. C. Johnston, "The ICH Score A Simple, Reliable Grading Scale for Intracerebral

Hemorrhage", Stroke **32**, 4, 891–897, URL http://stroke.ahajournals.org/content/32/4/891$\delimiter"026E30F$nhttp://stroke.ahajournals.org.ezp-prod1.hul.harvard.edu/content/32/4/891.abstract$\delimiter"026E30F$nhttp://stroke.ahajournals.org/content/32/4/891.full.pdf (2001).

Hirono, N., H. Kitagaki, H. Kazui, M. Hashimoto and E. Mori, "Impact of White Matter Changes on Clinical Manifestation of Alzheimer's Disease : A Quantitative Study", Stroke **31**, 9, 2182–2188, URL http://stroke.ahajournals.org/content/31/9/2182.long (2000).

Iliadis, G., V. Kotoula, A. Chatzisotiriou, D. Televantou, A. G. Eleftheraki, S. Lambaki, D. Misailidou, P. Selviaridis and G. Fountzilas, "Volumetric and MGMT parameters in glioblastoma patients: Survival analysis", BMC Cancer **12**, 1, 3 (2012).

Immerkæ r, J., "Fast Noise Variance Estimation", Computer Vision and Image Understanding **64**, 300–302 (1996).

Ivanovska, T., R. Laqua, L. Wang, A. Schenk, J. H. Yoon, K. Hegenscheid, H. Völzke and V. Liebscher, "An efficient level set method for simultaneous intensity inhomogeneity correction and segmentation of MR images", Computerized Medical Imaging and Graphics **48**, 9–20 (2016).

Kass, M., A. Witkin and D. Terzopoulos, "Snakes: Active contour models", INTERNATIONAL JOURNAL OF COMPUTER VISION **1**, 4, 321–331 (1988).

Kessler, L. G., H. X. Barnhart, A. J. Buckler, K. R. Choudhury, M. V. Kondratovich, A. Toledano, A. R. Guimaraes, R. Filice, Z. Zhang and D. C. Sullivan, "The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions.", Statistical Methods in Medical Research **24**, 1, 9–26, URL http://www.ncbi.nlm.nih.gov/pubmed/24919826 (2015).

Kosior, J. C., S. Idris, D. Dowlatshahi, M. Alzawahmah, M. Eesa, P. Sharma, S. Tymchuk, M. D. Hill, R. I. Aviv, R. Frayne and A. M. Demchuk, "Quantomo: validation of a computer-assisted methodology for the volumetric analysis of intracerebral haemorrhage", International Journal of Stroke **6**, 4, 302–305 (2011).

Kothari, R. U., T. Brott, J. P. Broderick, W. G. Barsan, L. R. Sauerbeck, M. Zuccarello and J. Khoury, "The ABCs of measuring intracerebral hemorrhage volumes", Stroke **27**, 8, 1304–1305, URL papers://490fc991-209f-415a-8b63-00fbb1df9aa5/Paper/p5097 (1996).

Lefohn, A. E., J. E. Cates and R. T. Whitaker, "Interactive , GPU-Based Level Sets for 3D Segmentation", LNCS 2878 Proc MICCAI pp. 564–572 (2003a).

Lefohn, A. E., J. M. Kniss, C. D. Hansen and R. T. Whitaker, "Interactive deformation and visualization of level set surfaces using graphics hardware", IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control **10**, 75–82 (2003b).

Lefohn, A. E., J. M. Kniss, C. D. Hansen and R. T. Whitaker, "A streaming narrow-band algorithm: interactive computation and visualization of level sets", IEEE Transactions on Visualization and Computer Graphics **10**, 422–33 (2004).

Malladi, R., J. Sethian and B. Vemuri, "Shape modeling with front propagation: a level set approach", Pattern Analysis and Machine Intelligence, IEEE Transactions on **17**, 2, 158 –175 (1995).

McInerney, T. and D. Terzopoulos, "Deformable models in medical image analysis: A survey", Medical Image Analysis **1**, 91–108 (1996).

Menze, B., M. Reyes and K. Van Leemput, "The Multimodal Brain TumorImage Segmentation Benchmark (BRATS).", IEEE transactions on medical imaging **PP**, 99, 1 (2014).

Mettler Jr, F. A. and M. J. Guiberteau, *Essentials of nuclear medicine imaging* (Elsevier Health Sciences, 2011).

Mitchell, J. R., S. J. Karlik, D. H. Lee, M. Eliasziw, G. P. Rice and A. Fenster, "The variability of manual and computer assisted quantification of multiple sclerosis lesion volumes.", Medical physics **23**, 85–97 (1996).

Mitchell, J. R., S. J. Karlik, D. H. Lee and A. Fenster, "Computer-assisted identification and quantification of multiple sclerosis lesions in MR imaging volumes in the brain.", J Magn Reson Imaging **4**, 2, 197–208 (1994).

Morgenstern, L. B., J. C. Hemphill, C. Anderson, K. Becker, J. P. Broderick, E. S. Connolly, S. M. Greenberg, J. N. Huang, R. L. MacDonald, S. R. Messé, P. H. Mitchell, M. Selim and R. J. Tamargo, "Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association.", Stroke; a journal of cerebral circulation **41**, 9, 2108–29, URL http://www.ncbi.nlm.nih.gov/pubmed/20651276 (2010).

Mortensen, E. N. and W. A. Barrett, "Interactive segmentation with intelligent scissors", Graphical Models and Image Processing **60**, 5, 349 – 384, URL http://www.sciencedirect.com/science/article/pii/S1077316998904804 (1998).

Neal, M. L., A. D. Trister, S. Ahn, A. Baldock, C. A. Bridge, L. Guyman, J. Lange, R. Sodt, T. Cloke, A. Lai, T. F. Cloughesy, M. M. Mrugala, J. K. Rockhill, R. C. Rockne and K. R. Swanson, "Response classification based on a minimal model of

glioblastoma growth is prognostic for clinical outcomes and distinguishes progression from pseudoprogression", Cancer Research **73**, 10, 2976–2986 (2013a).

Neal, M. L., A. D. Trister, T. Cloke, R. Sodt, S. Ahn, A. L. Baldock, C. A. Bridge, A. Lai, T. F. Cloughesy, M. M. Mrugala, J. K. Rockhill, R. C. Rockne and K. R. Swanson, "Discriminating Survival Outcomes in Patients with Glioblastoma Using a Simulation-Based, Patient-Specific Response Metric", PLoS ONE **8**, 1 (2013b).

NVIDIA, "NVIDIA CUDA Programming Guide", (2012).

Osher, S. and R. P. Fedkiw, "Level set methods: An overview and some recent results", Journal of Computational Physics **169**, 2, 463 – 502 (2001).

Osher, S. and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations", Journal of Computational Physics **79**, 1, 12 – 49, URL http://www.sciencedirect.com/science/article/pii/0021999188900022 (1988a).

Osher, S. and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations", Journal of Computational Physics **79**, 1, 12–49 (1988b).

Roberts, M., J. Packer, M. C. Sousa and J. R. Mitchell, "A work-efficient GPU algorithm for level set segmentation", in "High Performance Graphics (HPG '10)", pp. 123–132 (2010).

Roberts, M., M. Sousa and J. Mitchell, "Level set segmentation of volume data", URL https://www.google.com/patents/US9082191, uS Patent 9,082,191 (2015).

Rosset, A., L. Spadola and O. Ratib, "OsiriX: An open-source software for navigating in multidimensional DICOM images", Journal of Digital Imaging **17**, 205–216 (2004).

Rother, C., V. Kolmogorov and A. Blake, ""grabcut": interactive foreground extraction using iterated graph cuts", in "ACM SIGGRAPH 2004 Papers", SIGGRAPH '04, pp. 309–314 (ACM, New York, NY, USA, 2004), URL http://doi.acm.org/10.1145/1186562.1015720.

Saba, L., H. Gao, E. Raz, S. V. Sree, L. Mannelli, N. Tallapally, F. Molinari, P. P. Bassareo, U. R. Acharya, H. Poppert and J. S. Suri, "Semiautomated analysis of carotid artery wall thickness in MRI", Journal of Magnetic Resonance Imaging **39**, 6, 1457–1467 (2014).

Sacco, S., C. Marini, D. Toni, L. Olivieri and A. Carolei, "Incidence and 10-Year Survival of Intracerebral Hemorrhage in a Population-Based Registry", Stroke **40**, 2 (2009).

Saha, P. and J. Udupa, "Iterative relative fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation", in "Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on", pp. 28 –35 (2000).

Sanfilipo, M. P., R. H. B. Benedict, B. Weinstock-Guttman and R. Bakshi, "Gray and white matter brain atrophy and neuropsychological impairment in multiple sclerosis.", Neurology **66**, 5, 685–92, URL http://www.neurology.org/content/66/5/685 (2006).

Sethian, J. A., *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, vol. 3 (Cambridge university press, 1999).

Shortliffe, E. H. and J. J. Cimino, *Biomedical informatics* (Springer, 2006).

Styner, M., J. Lee, B. Chin and M. Chin, "3D segmentation in the clinic: A grand challenge II: MS lesion segmentation", MIDAS pp. 1–6 (2008).

T, B., J. Broderick, R. Kothari, W. Barsan, T. Tomsick, L. Sauerbeck, J. Spilker, J. Duldner and J. Khoury, "Early hemorrhage growth in patients with intracerebral hemorrhage.", Stroke. **28**, 1, 1–5 (1997).

Tsang, O., A. Gholipour, N. Kehtarnavaz, K. Gopinath, R. Briggs and I. Panahi, "Comparison of tissue segmentation algorithms in neuroimage analysis software tools.", Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference **2008**, 3924–3928 (2008).

Udupa, J. K. and S. Samarasekera, "Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation", CVGIP: Graphical Model and Image Processing **58**, 3, 246–261 (1996).

Wang, C.-W., C.-J. Juan, Y.-J. Liu, H.-H. Hsu, H.-S. Liu, C.-Y. Chen, C.-J. Hsueh, C.-P. Lo, H.-W. Kao and G.-S. Huang, "Volume-dependent overestimation of spontaneous intracerebral hematoma volume by the ABC/2 formula.", Acta radiologica (Stockholm, Sweden : 1987) **50**, 3, 306–11, URL http://www.ncbi.nlm.nih.gov/pubmed/19173095 (2009).

Webb, A. J. S., N. L. Ullman, T. C. Morgan, J. Muschelli, J. Kornbluth, I. A. Awad, S. Mayo, M. Rosenblum, W. Ziai, M. Zuccarrello, F. Aldrich, S. John, S. Harnof, G. Lopez, W. C. Broaddus, C. Wijman, P. Vespa, R. Bullock, S. J. Haines, S. Cruz-Flores, S. Tuhrim, M. D. Hill, R. Narayan and D. F. Hanley, "Accuracy of the ABC/2 Score for Intracerebral Hemorrhage: Systematic Review and Analysis of MISTIE, CLEAR-IVH, and CLEAR III", Stroke **46**, 9, 2470–2476 (2015).

Whitaker, R. T., "Volumetric Deformable Models: Active Blobs", in "Visualization In Biomedical Computing 1994", pp. 122–134 (1994).

Woo, D., J. P. Broderick, , N. I. o. N. D. Stroke, C. Sudlow, C. Warlow, C. Sudlow, C. Warlow, J. Broderick, T. Brott, T. Tomsick, et Al., J. Mohr, L. Caplan, J. Melski, et Al., J. Bamford, P. Sandercock, M. Dennis, et Al., J. Broderick, T. Brott, T. Tomsick, et Al., J. Broderick, S. Phillips, J. Whisnant, et Al., R. Fogelholm, M. Nuutila, A.-L. Vuorela, M. Giroud, E. Creisson, H. Fayolle, et Al., J. Broderick, T. Brott, J. Duldner, et Al., P. Daverat, J. Castel, J. Dartigues, et Al., R. Portenoy, R. Lipton, A. Berger, et Al., S. Tuhrim, J. Dambrosia, T. Price, et Al., S. Tuhrim, J. Dambrosia, T. Price, et Al., W. Young, K. Lee, M. Pessin, et Al., T. Brott, J. Broderick, R. Kothari, et Al., Y. Fujii, R. Tanaka, S. Takeuchi, et Al., S. Kazui, H. Naritomi, H. Yamamoto, et Al., K. Wagner, G. Xi, Y. Hua, et Al., G. Xi, R. Keep, J. Hoff, K. Lee, A. Betz, S. Kim, et Al., G. Xi, K. Wagner, R. Keep, et Al., J. Gebel, C. Sila, M. Sloan, et Al., J. Gebel, T. Brott, C. Sila, et Al., S. Greenberg, S. Finklestein, P. Schaefer, S. Greenberg, H. O'Donnell, P. Schaefer, et Al., G. Roob, R. Schmidt, P. Kapeller, et Al., H. Offenbacher, F. Fazekas, R. Schmidt, et Al., F. Cole, P. Tates, C. Fisher, R. R. Russell, S. Takebayashi, D. Woo, L. Sauerbeck, B. Kissela, et Al., H. Okazaki, J. Whisnant, H. Vinters, J. Vonsattel, R. Myers, E. Hedley-Whyte, et Al., T. Mandybur, S. Bates, K. Maruyama, S. Ikeda, T. Ishihara, et Al., K. Ueda, Y. Hasuo, Y. Kiyohara, et Al., S. Greenberg, J. Vonsattel, J. Stakes, et Al., M. Tomonaga, H. Vinters, J. Gilbert, I. Drury, J. Whisnant, W. Garraway, M. Hill, F. Silver, P. Austin, et Al., S. Greenberg, M. Briggs, B. Hyman, et Al., S. Greenberg, G. Rebeck, J. Vonsattel, et Al., S. Greenberg, J. Vonsattel, A. Segal, et Al., J. Nicoll, C. Burnett, S. Love, et Al., H. O'Donnell, J. Rosand, K. Knudsen, et Al., H. Bevan, K. Sharma, W. Bradley, J. Ruiz-Sandoval, C. Cantu, F. Barinagarrementeria, G. Toffol, J. Biller, H. Adams, X. Zhu, M. Chan, W. Poon, M. Sarwar, W. McCormick, Z. Hang, Y. Shi, Y. Wei, R. Brown, D. Wiebers, J. Torner, et Al., R. Brown, D. Wiebers, G. Forbes, et Al., R. Brown, D. Wiebers, G. Forbes, P. Crawford, C. West, D. Chadwick, et Al., C. Graf, G. Perret, J. Torner, Y. Itoyama, S. Uemura, Y. Ushio, et Al., R. Spetzler, R. Hargraves, P. McCormick, et Al., R. Spetzler, N. Martin, O. D. Curling, D. Kelly, A. Elster, et Al., P. Porter, R. Willinsky, W. Harper, et Al., D. Rigamonti, M. Hadley, B. Drayer, et Al., I. Mason, J. Aase, W. Orrison, et Al., H. Craig, M. Gunel, O. Cepeda, et Al., J. Dubovsky, J. Zabramski, J. Kurth, et Al., M. Gunel, I. Awad, J. Anson, et Al., M. Gunel, I. Awad, K. Finberg, et Al., M. Gunel, I. Awad, K. Finberg, et Al., T. Garner, O. D. Curling, D. Kelly, et Al., D. Kondziolka, P. Dempsey, L. Lunsford, G. Augustyn, J. Scott, E. Olson, et Al., C. Cammarata, J. Han, J. Haaga, et Al., G. Malik, J. Morgan, R. Boulos, et Al., L. Michels, J. Bentson, J. Winter, E. Olson, R. Gilmor, B. Richmond, Y. Saito, N. Kobayashi, G. Roman, M. Fisher, D. Perl, et Al., K. Kikuchi, M. Kowada, H. Shioya, et Al., O. Press, P. Ramsey, R. Djindjian, J. Grollmus, J. Hoff, C. Hodgson, H. Burchell, C. Good, R. Houdart, R. Djindjian,

M. Hurth, H. Plauchu, J. de Chadarevian, A. Bideau, et Al., H. Garland, S. Anning, C. Smith, L. Bartholomew, J. Cain, A. Bideau, G. Brunet, E. Heyer, et Al., A. Bideau, H. Plauchu, A. Jacquard, et Al., K. McAllister, M. Baldwin, A. Thukkani, et Al., K. McAllister, K. Grogg, D. Johnson, et Al., C. Shovlin, A. Guttmacher, E. Buscarini, et Al., M. Porteous, J. Burn, S. Proctor, M. Porteous, A. Curtis, O. Williams, et Al., R. Bird, J. Hammarsten, R. Marshall, et Al., A. Furlan, J. Whisnant, L. Elveback, A. Wintzen, H. de Jonge, E. Loeliger, et Al., G. Petty, L. Lennihan, J. Mohr, et Al., T. B. A. A. T. f. A. F. Investigators, G. Albers, D. Sherman, GD, J. Paulseth, P. Petersen, C. Kase, R. Robinson, R. Stein, et Al., J. Radberg, J. Olsson, C. Radberg, N. t.-P. S. S. Group, M. Pfleger, E. Hardee, C. Contant, et Al., T. Brott, K. Thalinger, V. Hertzberg, H. Okada, H. Horibe, O. Yoshiyuki, et Al., K. Okumura, K. Iseki, K. Wakugami, et Al., A. Segal, R. Chiu, P. Eggleston-Sexton, et Al., M. Caicoya, T. Rodriguez, C. Corrales, et Al., R. Monforte, R. Estruch, F. Graus, et Al. and J. Broderick, "Spontaneous intracerebral hemorrhage: epidemiology and clinical presentation", Neurosurgery Clinics **13**, 3, 265–279 (1990).

Xu, C. and J. Prince, "Snakes, shapes, and gradient vector flow", Image Processing, IEEE Transactions on **7**, 3, 359 –369 (1998).

Xu, X., X. Chen, J. Zhang, Y. Zheng, G. Sun, X. Yu and B. Xu, "Comparison of the tada formula with software slicer: Precise and low-cost method for volume assessment of intracerebral hematoma", Stroke **45**, 11, 3433–3435 (2014).

Xue, W., C. Zwart and J. R. Mitchell, "Intelligent initialization and interactivity: Optimizing level sets for t1-weighted white matter segmentation", in "MICCAI Interactive Medical Image Computing (IMIC) Workshop", (2014).

Yushkevich, P. a., J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability.", NeuroImage **31**, 3, 1116–28, URL http://www.ncbi.nlm.nih.gov/pubmed/16545965 (2006).