COMPARISON AND ASSESSMENT OF SEMIAUTOMATIC IMAGE SEGMENTATION IN COMPUTED TOMOGRAPHY SCANS OF THE KIDNEY

By

Courtenay Glisson

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Engineering

May, 2010

Nashville, Tennessee

Approved:

Dr. Robert L. Galloway

Dr. Michael I. Miga

ACKNOWLEDGEMENTS

I would like to express my thanks to Dr. Bob Galloway and Dr. Michael Miga for their advice and guidance on this project. I would also like to thank the graduate students of BML and SNARL for all of their help and guidance. I am grateful for the image data and open case data provided by Dr. Duke Herrell and Dr. Peter Clark of Vanderbilt Medical Center. This work was supported by an Image Guided Kidney Surgery Discovery Grant.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Motivation

According to the most recent report from the National Cancer Institute's Surveillance, Epidemiology and End Results (NCI SEER) program, over 54,000 new cases of kidney cancer were expected in 2008. This latest compilation of data also displays a thirty year trend of continual and significant growth in kidney cancer incidence. While such a trend is disturbing, improvements in imaging which have allowed earlier discovery of small renal masses have likely contributed to the number of diagnoses each year, steepening the curve. Evidence for the contribution of these earlier incidental discoveries exists in the mortality rate over the same thirty year span. Mortality associated with kidney cancer rose significantly in the period from 1975-1992, then fell in the period between 1992-2005. The SEER statistics demonstrate that 1)kidney cancer affects a large number of people each year, 2)the incidence of this disease is steadily rising, and 3)research and priority given to this disease and its treatment have already resulted in a drop in mortality, a trend that can continue as treatment becomes more targeted and less invasive[1].

Renal cell carcinoma, which accounts for the vast majority of kidney cancer, responds poorly to non-surgical interventions. As a result, the standard of care for kidney tumors has been established as surgical, with the first steady success found in radical nephrectomies. These interventions involve removal of the kidney in its entirety, along with surrounding fat and the associated adrenal gland[2]. As imaging modalities advanced, smaller tumors became

detectable, many being identified as incidental findings before any physical symptoms occurred. Surgeons began to adopt nephron-sparing approaches when tumor size and location allowed, and these partial nephrectomies became particularly important when the cancer was present bilaterally or kidney function was already compromised. Studies comparing radical to partial nephrectomies found that the two methods' outcomes with regard to cancer recurrence were statistically equivalent when lesions were small[3, 4]. Additional work suggested that radical nephrectomies were more likely to result in the development of chronic renal disease than nephron-sparing approaches when glomerular filtration rate was used as an indicator of post operative renal health[5].

During the period of shift from radical to partial nephrectomies, minimally invasive approaches were also evolving. Laparoscopic radical nephrectomy has been shown to demonstrate lower blood loss, shorter hospital stay, and faster recovery times typically associated with minimally invasive surgery while maintaining comparable efficacy to the open procedure[6]. While laparoscopic partial nephrectomies are performed in numerous hospitals, the procedure is technically challenging as the surgeon is forced to give up tactile cues, maneuverability, and easy access to cooling mechanisms which minimize the effects of ischemia while the kidney vasculature is clamped. At the same time the surgeon is required to find and follow a precise tumor margin which he or she must estimate from preoperative images. This set of difficulties tends to result in longer warm ischemia times for laparoscopic partial resections, and this ischemic time can have a negative effect on kidney function[7, 8].

Background

Given the trends towards nephron-sparing surgery and minimally invasive procedures, significant improvements in outcome could be seen by providing the surgeon with additional tools and information to accomplish the task as quickly and accurately as possible. Incorporation of image guidance provides this extra quantitative information, allowing closer margins and more spared healthy tissue while requiring less manipulation for visual inspection. Image guidance involves the calculation of a transformation which links image space (for example, locations in a CT volume) and physical space (locations recorded in the operating room), allowing the surgeon to quantitatively incorporate preoperative information into a procedure. In its most general form, image guided procedures require image data, localization of guiding landmarks of interest in both image and physical space, calculation of a registration transformation between the two spaces, and a way to display this integration of information to the surgeon[9]. While neurosurgical cases formed the base of early image guidance research due to the availability of rigidly defined, bony anatomy for registration [10-13], later work began to explore the possibilities for guidance in the soft tissues of the thoracic and abdominal cavities [14-16]. The non-rigid attachment of organs such as those in the abdominal cavity made extrinsic, preoperative fiducial markers impractical. Intrinsic points of interest were also difficult to localize in physical and image space. Both perioperative markers [17] and intraoperative image acquisition [18] improve the registration problem, but they also both remove the surgeon's ability to bring preoperative planning data into the procedure. One solution which removes the need for 1:1 correspondence between image space and physical space fiducial markers is the use of a surface based registration. In this scenario, physical points are collected over large areas of the organ. Surfaces have been generated in liver guidance systems using laser

range scanners [19]. Since an equivalent surface must be generated in image space to complete the registration, the organ borders must be delineated in image space as well as physical space. This segmentation of the organ of interest must be performed with care, as it becomes a limiting factor in the system's ability to register the two spaces.

The absolute, true tumor border in a preoperative image will be imperfectly localizable due to the limited spatial sampling which digital imaging provides. The gold standard is typically taken as the hand-drawn segmentation of an expert. While hand segmentation is accepted as the gold standard, it is also very time consuming, and interrater variability exists even among hand segmentations. Variable presentation of the multiple renal cell carcinoma subtypes contributes to this issue. Subtypes are classified based on their genetic characteristics and their appearance under microscopic observation, and these differences translate into macroscopic changes visible in CT, which is the most commonly utilized imaging modality for kidney tumors [20]. Cell type, changes in microvasculature, calcification, and areas of necrosis or hemorrhage all contribute to the final image. Tumors may vary widely in size and homogeneity; smooth homogeneity, patchy heterogeneity, and apparent central/peripheral divisions may be seen depending on the cancer subtype, as locally varying cell type and tissue architecture alter parenchymal densities. These multiple tissue densities become visible as intensity variations on CT, greatly complicating the landscape to be segmented. Certain subtypes may also be more likely to infiltrate vasculature, the collecting system, or perinephric fat. Kidney tumors may appear hyperdense, hypodense, or isodense when compared to normal kidney parenchyma. When contrast is injected, some tumors enhance and brighten on CT, and some enhancing tumors retain contrast longer than normal tissue. Additionally, tumors may

involve complicating factors such as large thrombi, which can extend into the large veins returning to the heart [20, 21].

Clearly, these factors contribute to difficulties in consistent and simple delineation of pathologic kidney borders. To speed the process and make decisions more objective, many algorithms have been developed to help automate segmentation, though the numerous ways in which tumors may present does represent a significant obstacle. As an integral part of an eventual, full-fledged image guidance protocol for partial nephrectomies, the characterization and understanding of the particular problems and idiosyncrasies associated with segmentation of the kidney is necessary. This work explores the ability of semi-automatic segmentation methods to produce a kidney surface useful for the registration step in an image guidance setting. The methods used will be drawn from the Insight Toolkit (ITK), a freely available package of image processing classes [22]. Their semi-automatic nature will allow minimal user input to be required, while the algorithm creates and evaluates the final segmentations. This effort will focus on those types of tumors which are most amenable to partial nephrectomies – that is, those which are relatively small parenchymal tumors displaying no obvious infiltration of vessels, collecting system, or perinephric fat. The work will attempt to implement an algorithm which functions well for various tumor intensities relative to normal tissue in multiple contrast states, and will observe and explore difficulties encountered which cause deviation from the gold standard.

CHAPTER II

METHODS

Whether the goal is to extract a particular region of interest or to simplify an image by classifying it into a number of regions, the goal of segmentation is to use information contained in an image to separate pixels into discrete groups. Information used can be limited to what the image contains, or additional information may be incorporated, possibly from other images or *a priori* knowledge about what the segmentation should look like. A number of methods have been devised to automatically or semi-automatically segment images using information such as intensity of the target pixel, intensities of neighboring pixels or other groups of pixels, or information from similar images. Methods may seek to classify pixels based on a threshold intensity value, finding the optimal threshold by calculating the current statistics of the segmented regions. Regions of interest may grow from a starting point based on connectivity and incorporate image information as the segmented area grows, with cessation determined by how different the next candidate pixels are from their neighbors. Alternatively, target image pixels may be assigned by comparison of their information with previously segmented training sets. In the absence of training sets, various areas can be iteratively assigned after an initial division is estimated, with each iteration's assignment using the statistics of the previous iteration's segmented groups. A fifth class of methods involves parameterization of initial estimated shapes, which are then allowed to deform and grow to the boundaries of a region of interest. In these cases internal shape constraints as well as external image information guide the model's growth [23].

While each of these methods can provide a segmentation, each also has strengths and weaknesses. Simple intensity thresholding is convenient, but it will not distinguish similar tissue intensity in separate locations, will not separate similar tissues which abut, and will not perform well when intensities in the target region are not homogeneous. Region growing incorporates the connectedness of each new candidate pixel to the previously segmented area, but may have similar problems with inhomogeneity, causing separate seed areas to fail to grow together, holes, jagged edges, or other anomalies. In a three-dimensional region growing algorithm, such defects can be seen even after the smoothing effects of dilation and erosion of the binary segmentation (Fig.1).

Figure 1. Three-dimensional region growing segmentation reprinted from Newman et al[24]. Left: Binary image after region growing algorithm. Right: After additional dilation and erosion filters of the segmented binary were applied to aid closure/filling of small holes and crevices.

Classifier methods utilizing statistics derived from segmented training sets bring in expert information and may help provide solid values which a region would be expected to display, but for a single anatomic segmentation in a complex image, the statistics of highly variable nontarget areas lumped into one region may cause difficulty. Methods which use an initial segmentation to

create regions and calculate statistics may suffer from the same problem, as the goal in kidney segmentation is to have a binary segmentation with only two classes – kidney and nonkidney. Allowing these statistical algorithms to create delineations where they logically exist, dividing all statistically disparate tissues, requires selection of the correct target region and elimination of the others after the initial segmentation. Due to the variable intensities introduced by kidney pathology and contrast agents, both the targeted kidney area and the non-targeted area are likely to consist of multiple segmented regions, which must be properly combined to yield an accurate binary segmentation. Again, if the target is connected to non-target regions which are statistically similar, an oversegmentation can occur. Parameterized models provide an opportunity for customization of the forces driving and slowing the model's growth, as well as those variables which affect the curve's allowable shape. While these methods also require an initialization step, their ability to draw out a single region while maintaining smooth curves, as well as their easily manipulated energy variables, make them an attractive candidate for segmentation of the kidney. The initial contour moves toward the target object's boundaries in an iterative process, constrained by internal energy variables such as the contour's stretch and bending while being attracted to the target object's edges by driving external energy forces such as intensity gradients [23]. These parameterized models can encounter problems, however, in situations where a curve must split into two or join with another curve. Such situations require combination or creation of a new equation, and involve the temporary creation of sharp corners in place of smooth, differentiable curves. A method known as level sets is able to provide the ability to develop a smooth curve with controllable variables while remaining equipped to handle these problems[25].

Level set methods handle the merging and splitting problems by avoiding an explicit parameterization of the evolving curve. This implicit handling of difficult curve shapes is the result of embedding the segmentation contour in a higher dimensional function, a 3-D function in the case of 2-D segmentations. The 3-D surface evolves over time, driven by a speed function which utilizes information in the image to be segmented, such as intensity gradient and curvature. The actual segmentation curve at a given time is a horizontal cross section of the higher level function, or a level set, for which the function's value is zero [25]. The ability of level sets to implicitly handle multiple contours while leaving the force for evolution open to manipulation makes it an attractive algorithm for semi-automatic segmentations. The incorporation of curvature into the speed of the evolving contour provides a function similar to the energy constraints in the parameterized models, avoiding oversegmentation at image 'weak points' despite their lack of intensity gradient barriers. Level sets have been used in a number of medical image segmentation problems, including brain, liver, and kidney.

Despite their success, level set methods are not without cost. Level sets' use of a higher dimensional function translates to higher computational costs, and its various parameters involved in developing a speed term to push or contain the contour must be customized if user interaction is to be minimized [26]. To reduce computation times, the higher level function may be defined in a narrow band around the current contour [27]. The parameters involved in driving the contour's evolution can vary greatly, but by utilizing one or more sources of information, they must push the contour towards the target region edges while restricting the curve from including other areas. Information from the target image may be used, such as intensity gradients or local contour curvature. Information from prior, known segmentations may direct growth based on shape, intensity, or curvatures seen in the training set [28, 29]. The algorithm

must be tailored to the task at hand if it is to become mostly automated, and a balance must be struck between preventing aberrant segmentations yet allowing for the variation which comes with pathology. Training an algorithm to expect a certain shape may prevent an over-segmentation when borders are not well delineated, but this type of constraint may also prevent correct segmentation of data which deviate from the training images (Fig. 2). With so many possible parameters involved, it is unsurprising that most algorithms remain semi-automated, requiring some user interaction or correction. In contrast to those algorithms which attempt to tailor results for an application, some work has focused on implementing efficient computation at the graphics processing unit level, leaving the user open to explore the parameter effects. This approach introduces flexibility, but also places responsibility on the subjective user [26].

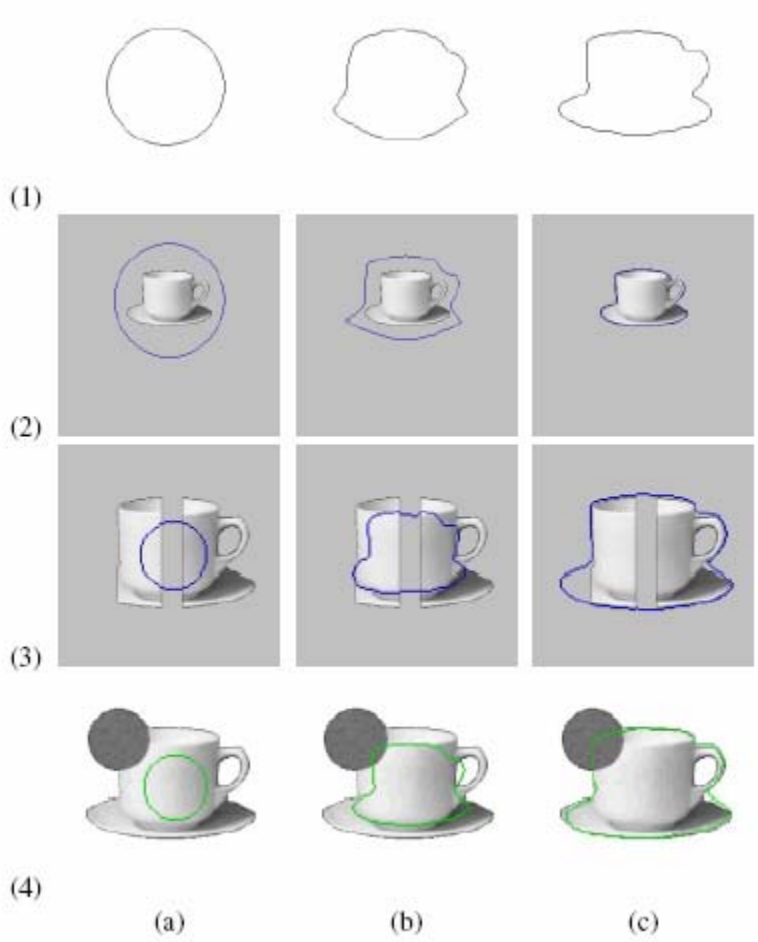

Figure 2. Level set segmentation using shape priors reprinted from Roussan et al[29]. Row (1) indicates the shape prior used as reference. Rows (2) and (3-4) show shape constrained geodesic active contours and regions, respectively. Shape constraints can infer missing information, but may fail to detect new shapes if they are dissimilar from the shape prior.

With the advantages and obstacles of level sets in view, this work focuses on exploring the ways in which several different level set implementations would behave when their parameters were varied for the specific purpose of kidney segmentation. One project goal was to understand the particular problems which the kidney presents, and to discover how the flexibility of the level set variables could affect results. Thus the work focuses on the analysis and optimization of various available level set implementations from the Insight Toolkit. The resultant algorithm of the project was to handle various renal carcinoma subtypes in multiple

stages of contrast. As a mostly independent level set class implementation was unlikely to be universally successful, given the diversity of data on which it would run, the multiple classes studied were incorporated into a single algorithm after each was explored individually. For the final algorithm to be considered successful, it would need to provide agreement with expert hand-segmentations, and would in the end be required to provide surfaces which could be used in image guidance registrations.

All of the classes used employ level sets for 2-D image slices. The following generalized theory of 2-D level set implementation should be fairly consistent in any algorithm [30]. A 3-D surface is defined as φ, a function of position and time, along with its zero level set Γ, the set of all points for which φ intersects the z-plane.

$$\Gamma = \left\{ \vec{x} \mid \varphi(\vec{x}, t) = 0 \right\} \tag{1}$$

The initial 3-D function created is typically the signed distance function from an initial estimate of the segmentation contour, and φ values at locations far from the contour itself are non-critical. The motion of the 2-D curve of interest (the segmentation contour) is observed as a velocity field deforms the 3-D surface over time, altering the surface's shape at the intersecting plane. These velocities reflect the desired movement of the contour intersecting any given point, and their contribution is controlled by the gradient of the surface at each point.

$$\frac{\partial \varphi}{\partial t} + \vec{v} \cdot \nabla \varphi = 0 \tag{2}$$

For the curve to expand or contract, only the component of velocity normal to the curve is desired. The level curves of a surface intersect its lines of steepest gradient descent perpendicularly, leading to the use of the gradient direction to determine velocity contribution. This relationship helps ensure that φ will not change drastically over a short timestep, as a steep

gradient attenuates the effect of the velocity term. Since the normal component is all that is needed, Equation (2) can be simplified to scalar quantities by the following:

$$\text{where} \quad v_N = \vec{v} \cdot \frac{\nabla \varphi}{|\nabla \varphi|} = |\vec{v}| \cos\theta,$$

$$\frac{\partial \varphi}{\partial t} + \vec{v} \cdot \nabla \varphi = \frac{\partial \varphi}{\partial t} + |\vec{v}||\nabla \varphi| \cos\theta = 0$$

$$\frac{\partial \varphi}{\partial t} + v_N |\nabla \varphi| = 0 \tag{3}$$

The above substitution allows a velocity term to be constructed while ensuring that it applies its force in the correct way. Using a signed distance function whose gradient is constant, a constant velocity field would imply a surface which moved only up or down, uniformly growing and shrinking the zero level set contour. The velocity values, however, generally incorporate information from the image (such as intensity or intensity gradients), as well as properties of the function itself such as curvature.

For each iteration of an algorithm, the height (φ) of every coordinate will change, and if velocity terms are non-homogeneous, then the resulting surface will have a different 3-D shape. The new segmentation contour is then found by locating the zero level set of that surface, after which the next iteration can begin. To simplify calculations, the new surface can be discarded in favor of a new signed distance function with respect to the latest zero level set contour, ensuring that gradients around the contour are not extreme.

The particular implementations used for this project consisted of multiple level set classes from the Insight Toolkit. Four classes were incorporated: FastMarching, CurveDetection, GeodesicActiveContour, and ThresholdLevelSets. Each method and its free parameters are briefly described below.

FastMarching – The fast marching method involves a special case of level sets, in which the propagation front is assumed to move only in one direction. Since the front can only cross each point once, the higher level function is built as the time crossings of an expanding contour driven by the speed function. The desired contour is found by specifying a timepoint and locating the corresponding contour. The ITK fast marching class requires both a time crossing value (at which to extract the contour) and a speed function. The speed function used was a processed gradient magnitude of the image to be segmented, and included two variables – k1 and k2 – which determined how steep the gradient must be to be considered an edge (mapped to a speed of zero), and what average gradient value represents the inside of the target anatomy (mapped to a speed of one). Higher k1 values require a larger gradient magnitude value for a pixel to be mapped to zero, making edges thinner and wispier in the speed image. Higher k2 values also tend to thin edges as higher gradient values in the kidney are mapped to one.

CurveDetection – The curve detection class employs the more generalized form of the level set method, and thus requires an initial level set (an initial 3D function). Like the FastMarching method, it requires a speed function, and also takes two scaling parameters – curvature and propagation scaling. Propagation scaling provides the relative amount of 'push' on the contour as it grows outward, and the curvature term is proportional to the contour's allowable radius of curvature at a given point. Higher curvature terms require higher radii of curvature, limiting sharp bends and smoothing the curve.

GeodesicActiveContour – The active contours method is similar to the curve detection method, involving an initial level set, speed function, propagation, and

curvature parameters. With the addition of an internally computed advection field, a final advection scaling parameter is introduced. Because the vector field created actually provides directional guidance towards edges, this method can effectively 'call back' slight initial oversegmentations. It is also less likely to stop at a thick edge's inner border than the curve detection method.

ThresholdLevelSets – The threshold method also involves an initial level set, as well as propagation and curvature scaling, but employs direct intensity information from the target image instead of using a separate, processed speed image. This algorithm requires a high and low intensity threshold to determine boundaries for the growing curve.

With these basic methods to use, five anonymized CT image sets in DICOM format were obtained with which to optimize an algorithm. Pathological and normal kidneys in various phases of contrast were part of the ten kidney set. Contrast phase was classified according to the following: non-contrast – no cortico-medullary delineation visible; cortical contrast – early/arterial phase, cortex brighter than medulla with arterial contrast; corticomedullary contrast – mid-phase, cortex roughly matching medulla with minimal venous washout; and medullary contrast – late phase, medulla brighter than cortex with relatively higher venous intensity. Hand segmentation was performed on each image set and was considered the "gold standard" for all comparisons to the automatic methods.

The first requirement in building the algorithm was to choose a metric for comparing auto segmentations to the gold standard hand segmentations. For each CT image/segmentation method combination, the algorithm was run by hand, choosing seeds and parameters which seemed to perform best under visual inspection. These auto segmentations were compared to

their hand segmented counterparts using a kappa statistic, a second, related metric called a similarity measure, and the more simple percent overlap between the two images. The kappa statistic compares the agreement observed between two raters of a set of binary decisions with the agreement expected if those two raters' assessments were entirely independent [31]. The division of pixels by hand and auto segmentations are shown below along with their use in the kappa statistic calculation.

	AutoYes	AutoNo	
HandYes	a	b	= a+b
HandNo	c	d	= c+d
	=a+c	=b+d	

$$p_{obs} = \frac{(a+d)}{n},\ p_{\exp} = \frac{(a+c)(a+b)+(b+d)(c+d)}{n^2}$$

$$k = \frac{(p_{obs} - p_{\exp})}{1 - p_{\exp}} \tag{4}$$

In each CT/method combination, results showed that the kappa statistic could be influenced easily by including the large area of the CT which both raters left unsegmented (greatly increasing the amount of agreement between auto and hand segmentations). Leaving this area out entirely negatively skewed the statistic, as the expected unsegmented agreement calculated was always higher than the observed agreement of zero. The similarity measure metric, in contrast, only involves a, c, and b from the rater's square above, making the assumption that black space in the image is very large compared to the target area [32]. After substituting for the observed and expected percentages, this assumption allows simplification of the similarity equation as seen in Equation (5).

$$k = \frac{2(ad - bc)}{(a+b)(b+d)+(a+c)(c+d)}$$

$$S = \frac{2a(INF)}{(a+b)(INF)+(a+c)(INF)}$$

$$S = \frac{2a}{(a+b)+(a+c)} \tag{5}$$

This metric yielded results which followed a similar trend to percent overlap in all cases. The similarity measure accounts for the intuitive conclusion that one rater's segmentation contained within a second rater's boundaries is a more accurate segmentation than two areas which overlap but whose centroids are farther apart [32]. As the similarity metric emphasizes location as well as the amount of overlap and does not include the large area which both raters agree is not part of the target, this measure was chosen for all further comparisons between segmentations.

With a suitable metric for comparison chosen, a sensitivity test was run on the only required user input which would not be automated in the algorithm – the seed points. Each method had an estimation of 'tailored' seed points created during the previous step, which were chosen to visually help the method succeed. Each method/CT combination was then run again with four additional seed sets – the three sets tailored to the other methods, and an 'ideal' set which was chosen without regard to a given method's success. Similarity values were calculated both on a per slice and whole image basis. Analysis of the similarity measure over the depth of the kidney for all seed sets used on a given image/segmentation method combination showed that while some variation existed among the seedsets, this variation was typically small compared to the variation in similarity seen by changing the method used to segment. In addition, even these small differences failed to consistently show that a method improved with the use of its tailored seeds, leading to the conclusion that an algorithm could be built from all four methods while

remaining robust to an individual user's seed choice. The 'ideal' seed set would therefore be used for all further auto segmentations.

For an algorithm using the chosen ITK classes to be constructed, numerous variables required optimization. Initially, each method was optimized alone. In this step, the methods were run multiple times, varying each free parameter individually. Both overall (for entire image sets) and slice-by-slice similarity values were calculated. All image sets were analyzed as transverse slices, with axial depth varying along the superior-inferior direction. The trends seen over the axial depth of the kidney led to the use of depth-specific values (values corresponding to the poles or to central slices), and the observation that contrast had a significant effect on results led to the creation of contrast state as an input variable to the algorithm, as this state is easily discernable in viewing an image set.

The FastMarching method's input variables were the k1 and k2 values used in the speed function (created from the gradient magnitude of the image to be segmented). Seven pairs of k values were chosen. Similarity trends were analyzed for the pairs of k values for each CT. Values for k1 which were too low were found to create barriers to contour growth, particularly when high gradients were present. Fairly high k1 values were tolerable when k2 was kept low, with mid-value k1 and low k2 yielding good results for two of the images which lacked contrast agent ("non-contrast" images). With three non-contrast images in the testing set, variation could be observed in kidney pathology, and not all non-contrast images were alike. High gradient non-contrast images (in this case, dark, fluid-filled, cystic tumors surrounded by normal intensity parenchyma) appeared to improve in similarity when k values were used which thinned the borders of the speed image. Values of k2 which were higher and closer to k1 provided this thinning of borders, and unsurprisingly also resulted in better similarity values when high

gradients due to contrast were present. As gradients near the poles of the kidneys tended to be weaker, thicker borders were beneficial there. Since good similarity results appeared to depend heavily on the gradients present in the image, and these gradients were influenced by contrast and by partial volume effects near the kidney poles, k values were chosen for FastMarching which would be set automatically depending both on contrast state and the current axial depth in the kidney segmentation. Knowing that some failures would be inevitable, and having observed the difficulty in choosing a one-size-fits-all value for the non-contrast state, failure options were outlined which could tighten or relax the speed function by thinning or thickening its borders.

The CurveDetection method required the k-value parameters of the previous method, and also incorporated curvature and propagation scaling terms. Five k-value pairs were chosen, and three runs were performed for each pair. Since only the ratio of curvature to propagation was important, curvature was set and the three tests were run with propagation at 10x, 100x, and 1000x the curvature setting. Again, two of the three non-contrast images yielded similar results, while the third with its higher image gradients performed better under different parameters. In general, non-contrast or lower gradient images seemed to require a tighter speed function, and the speed function terms tended to dominate the results over changes in propagation. With a secure speed function, increases in propagation made relatively little difference, while with a loose speed image lower propagation values with respect to curvature did not rein in bleeds. The segmentation of higher gradient images could be impeded by a tight speed image, and performed better when only the highest gradients in the image were assigned a zero value. While increasing the propagation term in too-tight speed images provided minimal aid in some cases, the speed image values still dominated the outcomes of the segmentations. At kidney ends, tighter speed functions lifted similarity metrics in some cases. Based on these observations, algorithm

parameters for CurveDetection were set to vary with contrast state and depth, and options were outlined which could tighten or relax the speed function in case of failure.

The GeodesicActiveContour method was optimized next, using the previous methods' speed image parameters, curvature, and propagation terms, and adding an advection term. Again, the importance of the curvature, propagation, and advection terms lay in their ratios. As the number of test runs increased with each additional parameter, only three k-value pairs were chosen to test this method. Curvature was again set, and propagation and advection were varied. Runs were performed with propagations of 100x, 1000x, and 10000x the curvature value, and for each of these, the advection value was tested at 1x and 10x the propagation value. With the addition of the advection term, much more variation was observed for a given speed image. Two trends were observed for a given speed image, corresponding to advection on par and advection greater than propagation, regardless of curvature value. This reinforces the previous conclusion for the CurveDetection method that the propagation term changes did not have a great effect on outcome, while showing that advection could alter outcomes significantly. In most of the tests, the general trends showed that in the middle kidney sections, an advection term on par with the propagation term provided the best results, with the kidney ends benefiting from the higher advection term relative to propagation. Cases which required the higher relative advection term tended to be those with looser speed function terms, indicating that if borders were weakened, a gradient field pointing towards those borders could prevent segmentation bleeds. Speed image and relative scaling term values were again chosen based on depth location in the kidney as well as contrast state, and failure parameters were outlined.

ThresholdLevelSets, the final method, did not require a speed image, using instead a maximum and minimum intensity threshold along with curvature and propagation scaling terms.

As the CT numbers of fat and kidney parenchyma in various contrast states should remain fairly consistent, the threshold values for each contrast state were chosen by visual inspection, and test runs were only performed with varying propagation terms relative to a set curvature term. Propagations used were 0.01x, 0.05x, 0.1x, 0.5x, 1x, and 10x of the curvature term. In most images, the higher propagations tended to perform more consistently. With contrast, however, some failures could be recovered by low propagation terms. Parameters were chosen for the algorithm and failure values were outlined. This method was observed to run more quickly than the previous three, due mostly to the lack of need to build a gradient magnitude image.

Since each method could in some cases be significantly improved by parameter changes (Fig. 3), the final step in the optimization of the individual methods was to incorporate these changes upon detection of a failure. No one set of parameters would hold for all circumstances, so a correctly identified failure would yield the opportunity to rein in over-segmentations with tighter parameters, or alternatively to lift restrictions to contour expansion in the case of a small or disconnected segmentation.

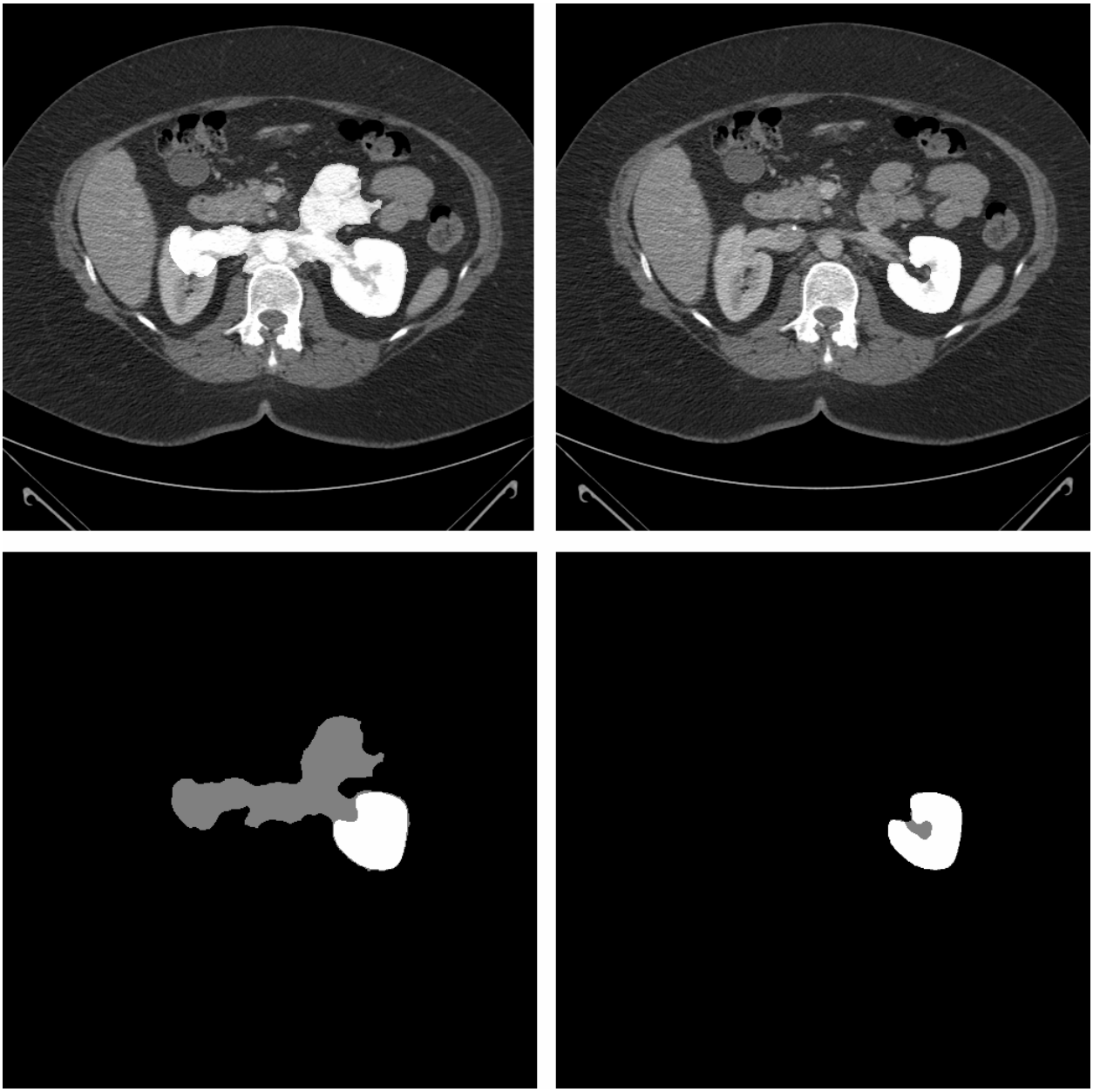

Figure 3. Semi-automatic algorithm corrected by parameter change. Left column: ThresholdLevelSet method using high propagation scaling causes an oversegmentation due to easily followed vasculature on CT. Lower panel shows the extent of the oversegmentation overlaid onto the handsegmentation. The white area represents pixels segmented by both the handsegmenter and the algorithm. Right column: The same method's results using lower propagation scaling, automatically implemented upon detection of the preceding failure. Lower panel again shows autosegmentation overlaid onto handsegmentation, with the white area representing pixels chosen by both methods. The autosegmentation undersegments the calyceal area, leaving the gray defect.

Identifying a failed segmentation in a 2D pathologic kidney presents numerous problems. Kidneys vary in natural size and shape, and pathologic conditions can mean that the 'correct' shape of a segmentation may be considerably different from that expected when viewing a normal kidney. Several basic types of failures could be identified, however. If the chosen seedpoints failed to grow together, a disconnected segmentation could be identified. This situation would very likely represent a failure, as a normal kidney slice will be visible as one contiguous target area on axial CT, and tumors are rarely extended enough to appear as a discrete area separate from the parenchyma. Kidneys with such extensive and invasive pathology would likely not represent the desired target population of this work, which focuses on partial nephrectomies, and so disconnects were incorporated as a detected failure. The size of the segmentation also provided clues. The overall in-slice height and width of a segmentation should be within reasonable bounds if the algorithm does not fail to grow or bleed into other areas of the CT image. To incorporate this information, the height and width of bounding boxes were found for each slice of all handsegmented kidneys. The slices were binned based on their depth within the kidney, and values within these bins were averaged. These average values, with some chosen error, represented the typical bounding box size of a mix of normal and pathologic kidneys.

Despite the irregularities introduced by pathology, a third failure recognition method was incorporated which drew more specifically on what a 'typical' segmentation might look like. All handsegmented kidneys, pathologic and normal, were used in this process, which layered the handsegmentation slices within a given axial depth range, creating a general probability map of location. All new additions to the atlas were layered by aligning their centroids and rotating the binary segmentations' axes of least inertia to lie along a common line. Each atlas bin was then

normalized by the number of images which contributed to it. This atlas was used for comparison to automatic segmentations when running the algorithm. When an autosegmentation result was placed over the correct atlas bin's composite probability image, the atlas's likelihood of that pixel being segmented was used to designate whether the image pixel agreed with the atlas or not. The same similarity measure used previously was then calculated. With this failure option, the atlas could be made more or less rigid (require a higher or lower probability of a pixel outcome to declare the autosegmentation a match to the atlas), and the final similarity measure could also be raised or lowered to alter how close an autosegmentation must be to the atlas for it to be considered a success.

Finally, with each method's parameters chosen, failure detection in place, and changes to make upon failure determined for individual methods, the algorithm's order of methods could be evaluated. While parameter changes could recover failures in some cases, the varying characteristics of each of the four methods led to some cases requiring a specific algorithm to succeed. A single method might oversegment using one set of parameters and then proceed to undersegment when using more rigid values, or alternatively it might not be able to rein in oversegmentations no matter how rigid its parameters (Fig. 4).

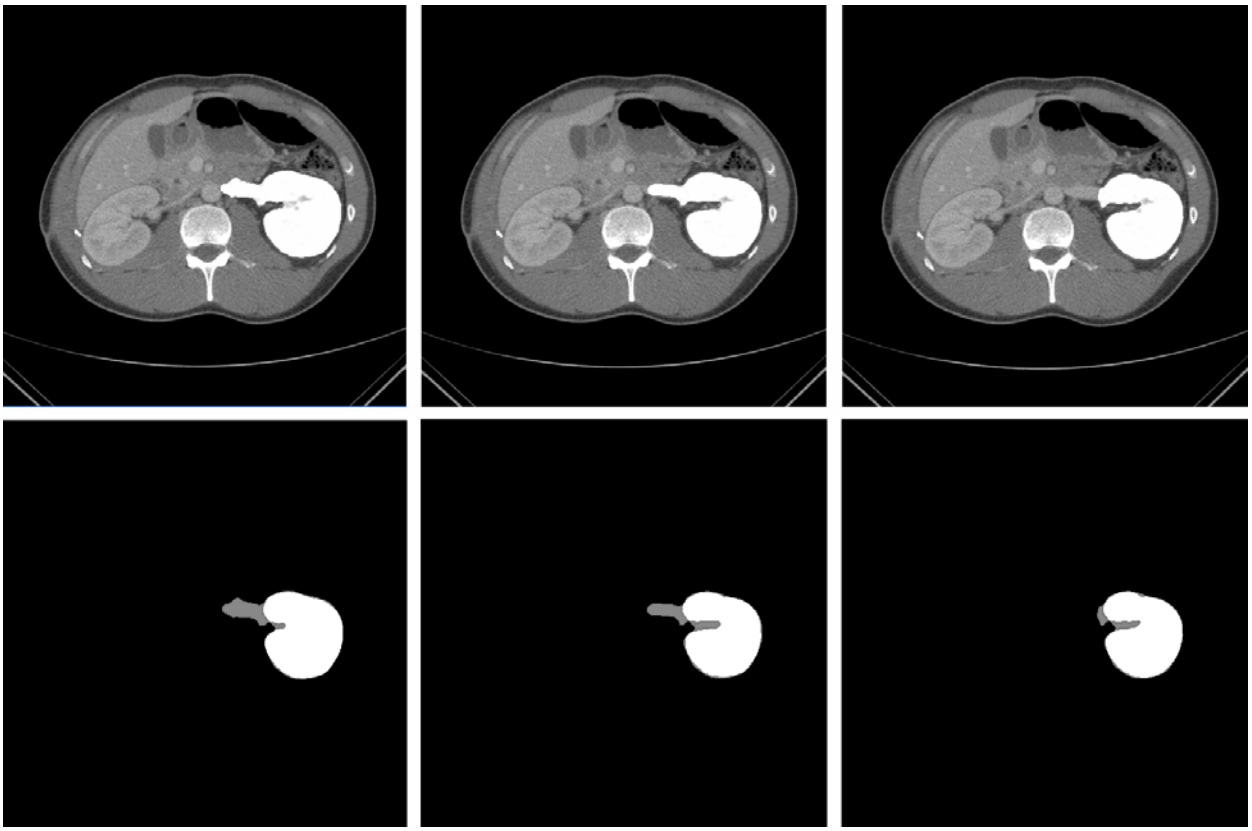

Figure 4. Semi-automatic algorithm corrected by method change. Left column: ThresholdLevelSet method using low propagation scaling cannot contain an oversegmentation caused by vessel following. Lower panel shows the extent of the oversegmentation overlaid onto the handsegmentation. The white area represents pixels segmented by both the handsegmenter and the algorithm. Middle column: CurveDetection method's optimal results, showing only slight improvement in avoiding the vessel bleed. Lower panel again shows autosegmentation overlaid onto handsegmentation. Right column: GeodesicActiveContour method results, demonstrating restriction of the vessel oversegmentation.

To choose the most efficient order for implementation of the optimized methods, each method was run on each testing CT image, allowing the algorithm to determine failure and alter parameters as needed. The method similarity values were observed for whole kidneys as well as for central and polar (inferior/superior end) kidney sections. Average similarity values of successful slices were compared directly, and percent failure values were also calculated based on the number of slices in a section with $S<0.8$. These results gave an indication of how well a

method performed when it was successful, as well as how likely it was to fail. Using the similarity values, the percentage of successful slices segmented, and other factors such as the speed of the method, the order of the optimized methods was set for the algorithm. For the kidney poles, the GeodesicActiveContour method yielded the most successful results. For middle slices, the ActiveContour and CurveDetection methods showed the fewest failed slices, but these methods were placed behind ThresholdLevelSets in the queue because when the latter method does work, it is significantly faster.

CHAPTER III

RESULTS

The final segmentation algorithm operated semi-automatically, requiring only seedpoints and the image's contrast state. The algorithm was run on all of the kidneys used in its training. The whole kidney comparisons below show an average similarity of 0.93, and average values for each of the contrast states remains above 0.90. Two values are shown for each CT, as the right and left kidney for each was segmented separately.

Table 1. Training Kidney Results – by Contrast

	NoContrast	CContrast	MContrast
CT1	0.914		
	0.927		
CT2	0.935		
	0.916		
CT3		0.952	
		0.946	
CT4			0.965
			0.948
CT5	0.907		
	0.894		
Average	0.916	0.949	0.957
StDev	0.014	0.004	0.012

Table 2. Training Kidney Results – Total

	All Kidneys
Average	0.930
StDev	0.023

The training kidney results indicate that the algorithm performed well on images for which it was tailored. An evaluation group of kidneys (7 image sets, 14 kidneys) segmented to determine if

the algorithm would continue to perform acceptably on new data. These results are shown below. The average similarity value for all kidneys is again 0.93, with all contrast state grouped averages above 0.90

Table 3. Evaluation Kidney Results – by Contrast

	NoContrast	CContrast	MContrast
CT6		0.928	
		0.909	
CT7		0.938	
		0.930	
CT8		0.938	
		0.945	
CT9			0.937
			0.935
CT10	0.920		
	0.946		
CT11			0.953
			0.912
CT12			0.930
			0.866
Average	0.933	0.932	0.922
StDev	0.018	0.013	0.031

Table 4. Evaluation Kidney Results – Total

	All Kidneys
Average	0.928
StDev	0.022

To ensure that the algorithm was capable of delivering acceptably reproducible results, three of the evaluation image sets were hand segmented by a second rater, and a second seedset was chosen for the auto segmentation by the same second rater. Similarity values were calculated comparing the two hand segmentations to each other and comparing the two auto segmentations to each other. A paired t-test compared the interrater handsegmentation similarity to the interrater autosegmentation similarity, indicating that within a 95% confidence interval,

the automatic segmentation showed less variability between raters than did the hand segmentation.

These results indicate that the algorithm performed acceptably with respect to the gold standard of hand segmentation. The final test of a segmentation algorithm for image guidance, however, is evaluation in a practical setting. Three open, partial nephrectomy cases were observed, and laser range scans were taken during each. These scans served as point cloud data to be registered to a segmented CT surface. Both hand segmentations and auto segmentations of the CT images were performed, and the perioperative laser range scans were registered to each. An initial point-based registration was performed to provide an initial alignment. The iterative closest point method of surface-based registration was performed for final alignment. Mean and maximum closest point distances between the two surfaces for each case were calculated (Table 5).

Table 5. Closest point differences between registered laser range scan data and segmented CT surfaces (all distances in mm).

Case1	Max	Mean	StDev
Handsegmentation	5.477	0.874	0.781
Autosegmentation	6.613	0.819	0.808
Case2			
Handsegmentation	9.120	0.917	1.023
Autosegmentation	9.428	1.449	1.673
Case3			
Handsegmentation	4.372	0.906	0.714
Autosegmentation	3.695	0.808	0.629

These distance data show that the autosegmentation is capable of producing similar registration results to the gold standard of handsegmentation, with mean closest point distances of approximately 1mm. Visual inspection of the registered surfaces confirmed that the automatic

method's registrations were comparable to those using handsegmentation (Figs. 3). This level of surface matching provides support for the incorporation of the semi-automatic method into an image guidance system.

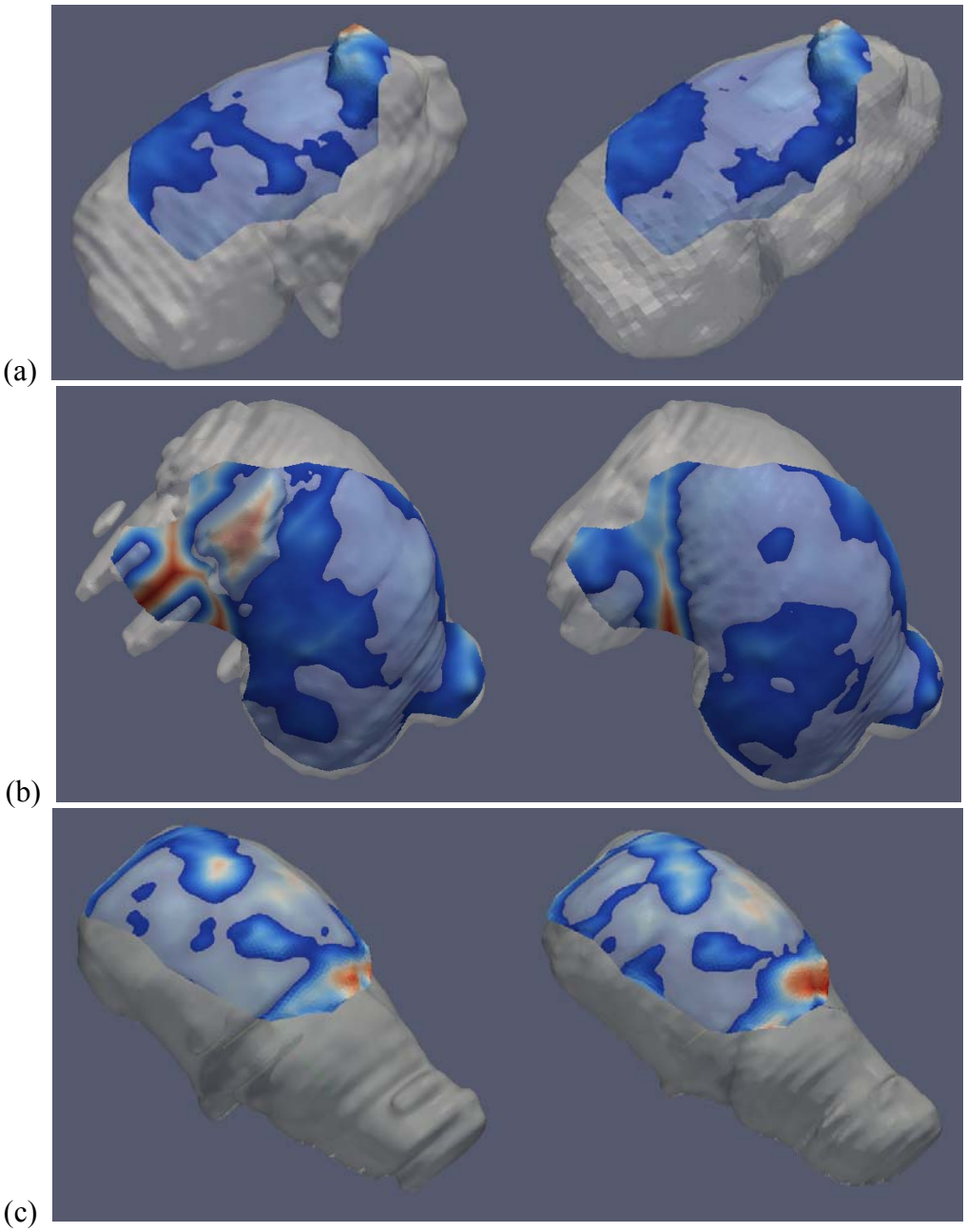

Figure 5. Preoperative CT and intraoperative laser range scans for three cases, (a-c). Registered laser range scan shown with autosegmented (left) and handsegmented right) CT surfaces. Laser surface colors indicate closest point distances to the CT surface, with the highest distances visible as red.

The fits seen for case (a) are similar for hand and autosegmentations, with the highest closest point distances seen over the tumor itself. This area of higher error is expected, as the surgeon generally leaves an area of perinephric fat undisturbed in the area of the tumor, while the CT segmentations include no fat cap. Case (b) was included despite lying outside of the previously stated scope of the algorithm, which was to target smaller, exophytic masses which would be most amenable to partial nephrectomy. This case also displayed a good visual and numerical fit over the CT image, though higher areas of error exist for the autosegmentation. The failure of the autosegmentation in this large tumor is unsurprising, as the error detection methods of the algorithm rely on the shapes seen in its training set, none of which were so large as in this case. For this tumor, successful segmentations would be likely labeled as too large and dissimilar from the atlas, and the algorithm would classify such a case as a failed attempt. Both cases highlight the role of unambiguous tumor shapes in the correct alignment of the final registration, while sufficient normal parenchymal surface provides weighting to bring the surfaces together. In this way, the surfaces can be reliably brought to the correct closely aligned position, instead of settling into an incorrect minimum, and the normal parenchymal areas are more closely aligned than areas such as the fat cap over the tumor. Case (c) displays a more subtle shape due to a more internally located tumor. While the tumor shape is still important in determining the correct final alignment, this case highlights the importance of a reasonable initial alignment of the data to be registered. The central area of high error for this case, visible for both hand and autosegmentations, may be the result of an artificially jagged CT surface, which was noted during handsegmentation. This artifact may be due to breathing during CT acquisition.

The mean closest point distances for each of the three cases are approximately 1mm, and the intraoperative data display good visual alignment to both hand and autosegmentation surfaces. While these data are encouraging, and each transformed laser range scan both quantitatively and qualitatively fits its respective segmentation surface, a further step was taken to determine the variability introduced into the calculated transformation by registration to autosegmentation versus handsegmentation. Since the same initially aligned laser range scan was registered using the same method to different segmentation surfaces, any differences in the resulting transformations were due to the use of one segmentation versus the other. The effect of this variability of transformation is important not only where the laser range data lies but also at subsurface and distal kidney points, and so the inverse of each transformation was applied to points in image space, and the distance between these transformed points was calculated. As handsegmentation is the gold standard, this calculation was performed for all points in the handsegmented CT image. The kidney was subdivided into five categories, with each category increasing in distance from the tumor centroid (Fig. 6). This type of grouping was chosen to discern any difference in transform variability when points were near or far from the surgeon's intraoperative target.

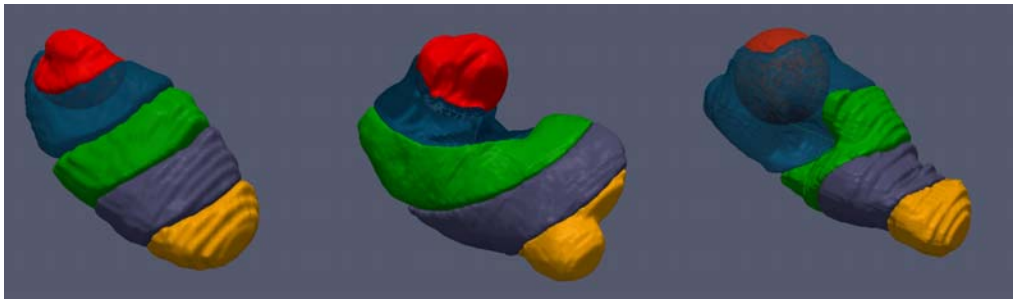

Figure 6. Visualization of distance grouping categories used in evaluation of the variability of transformation introduced by using handsegmentation versus autosegmentation. Points were classified into five categories by distance from the tumor centroid. Category (1) is shown as red, with distance increasing with each successive section.

For all three cases, the average variability in physical point location introduced by using one transformation versus the other was no more than a few millimeters (Table 6). Case (a) displays a slight trend, with lower variability near the tumor, but the same trend is not apparent in the other cases. Case (b) has two tumors providing descriptive surface information for the registration algorithm, but the gross difference in the autosegmentation surface at the larger, tumor makes a higher variability in this area unsurprising. The third case, with less obvious features to drive the two registrations to match in a particular location, displays the most consistent variability averages across categories.

Table 6. Average distances between resultant physical space points when both hand and autosegmentation resultant transformations are applied to image space coordinates, divided by category. Each category represents a distance grouping from the tumor's centroid, with category (1) being the closest (all distances in mm).

Distance Category	Case (a)		Case (b)		Case (c)	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
1	1.346	0.422	2.795	0.811	2.517	0.065
2	1.282	0.654	1.674	0.882	2.550	0.121
3	1.801	0.692	2.314	0.848	2.473	0.098
4	2.498	0.745	2.464	0.994	2.439	0.068
5	3.095	0.651	2.130	0.707	2.371	0.047

The results of these cases indicate that when used within its intended scope, the algorithm not only functions acceptably by direct comparison to the gold standard of handsegmentation, but that it is capable of producing a workable surface which may be used for registration to intraoperative data during the guidance process.

CHAPTER IV

DISCUSSION

The ability of a segmentation algorithm to correctly define kidney anatomy is essential to the success of surgical image guidance. Inaccurate segmentation limits registration accuracy, and thus the transformation used to guide surgical tools during an intervention. Issues which arise during the creation of a semi-automatic segmentation algorithm may be related to physiology, to the imaging process, or to the type and implementation of algorithm used. Biological considerations may include the proximity of the kidney to other organs, the difficulty in defining proper organ delineation where vessels enter the kidney, and the variety of tumor types seen in the kidney. The kidney may abut other organs, creating a border which is difficult even for expert hand segmenters. The calyceal area is visually complex, with the kidney parenchyma folding into itself at the vessel insertion, and defining borders between parenchyma and vessels or ureters can introduce additional variability even to the gold standard. The tissue seen in kidney tumors without contrast may vary, appearing dark or as almost indistinguishable from normal parenchyma. Furthermore, tumors may take up more, less, or equal amounts of contrast when compared to normal tissue. The considerations of the imaging process may relate to biology, as CT numbers are based on unchanging physiology, which means that the only changes which can be introduced in the imaging process to highlight the kidney are direct physical interventions such as contrast agent. Imaging limitations such as the partial volume effects seen at the kidney poles also create weaker image borders. Layered over the biology and imaging considerations are those of the algorithm's process itself. Several of the level set

implementations require speed images, which are built from gradient magnitude images. The gradient magnitude reflects information already present, and a hole in an organ contour cannot be closed if no gradient exists. Allowing smaller gradients to form contour borders may create or emphasize borders elsewhere, such as that between the cortex and medulla in contrast images, which in turn hinder the progress of the segmentation's growth. The algorithm's use of curvature can aid in holding back a segmentation when a small hole in a gradient image exists, but requiring a segmentation to be too smooth can also inhibit proper growth, as small ball-shaped tumors on the kidney's parenchyma do require relatively high curvature in the segmentation contour.

The issues associated with all three of these sources became apparent in building the autosegmentation algorithm. Weak organ borders caused by partial volume effects as well as those seen when the kidney abutted another organ required stronger speed image contours, as well as curvature parameter adjustments. Lower attenuation tumors created parenchyma-tumor borders, and so in these cases speed images could not be made too strong. Partial volume effects seen at kidney poles again required a stronger speed image. Since certain cases presented repeatable problems with relatively predictable solutions, parameters were set for those circumstances which could be easily recognized (kidney poles identified by percent depth in the kidney axially, contrast state given as an input variable), and automatic parameter changes were implemented for failures which appeared to be the result of parameters being too tight or too loose.

After attempting to understand and address the difficulties that biology, the imaging process, and the algorithm itself presented, the next requirement was to recognize a failed segmentation so that appropriate changes could be made to improve the likelihood of success. A

general size constraint was applied for segmentations of a particular axial depth, which was itself built from the average size of multiple kidneys in that depth category. To ensure that the size constraint was not incorrectly bypassed by multiple starter contours which failed to grow together, a connectedness requirement was also created. A more in depth method of comparing typical kidney size and shape to a segmented slice was implemented by creating a simple atlas of probability from the training kidneys, which could be compared to the segmented slice in place of a handsegmented kidney.

The final algorithm was run for all of its training kidneys and found to have a similarity metric of 0.93 when compared to the gold standard of handsegmentation, showing that the algorithm was built successfully. It was then run for several test kidneys, which were not used in its creation, which resulted in the same similarity metric. This consistency showed that the algorithm could function with data to which it had not been tailored. Interrater similarities for two handsegmentation raters and two autosegmentation raters were compared, revealing lower variability for the autosegmentation, and confirming that the algorithm was not particularly sensitive to the user's seed choices. The time required by the algorithm for segmentation varied with the difficulty of the segmentation (failed slices were re-segmented). For an average-sized kidney of forty slices, a more difficult segmentation was observed to take sixteen minutes, while a less difficult segmentation was completed in six minutes. Handsegmentation of the same volumes would take approximately twenty-five minutes. Additionally, the interaction time (seed selection process) required of the algorithm was less than five minutes regardless of difficulty. Based on these data, the algorithm appeared to compare favorably to the gold standard of handsegmentation in terms of time, work required, and resultant data.

Final validation of the algorithm was shown via the practical test of evaluating autosegmentations by methods used in actual surgical guidance. Autosegmentations were registered to intraoperative data and closest point distances were calculated, revealing that the autosegmentation yielded very similar closest point values to handsegmentation. Highest areas of error in one case were over the tumor, which was expected due to surgeon preference for leaving a cap of fat over the tumor to ensure positive tumor margins outside the kidney. Highest error areas for a second case reflected the algorithm's inability to perform well when presented with pathology too different from its training set. In this case, the parenchyma-tumor groove of this tumor in addition to a second, smaller tumor did provide unambiguous surfaces for the registration, and good visual alignment as well as low closest point distances for the smaller tumor were observed. A third case displayed similarly low registration error despite more subtle external tumor topology. Further exploration into the differences introduced by using one segmentation versus the other showed that the transform variability (the distance between the two physical space points found by applying both inverse transformations to a single image space point) was typically less than three millimeters. These results showed that the laser range scans were settling onto the same features during their registration to hand and autosegmentation surfaces, and that these features were similar in both segmentations. Each resultant transformed dataset not only matched its respective CT surface well, but also the other transformed dataset. Overall, the mean closest point distances for both methods of segmentation as well as the variability introduced by using one segmentation method versus the other were low, and the performance of the algorithm on those pathologies within its scope supports the assertion that the semi-automatic method functions comparably to the handsegmentation both when directly compared and when used as part of the image guidance workflow.

CHAPTER V

CONCLUSION

Using level set implementations found in the Insight Toolkit, an algorithm for semi-automatic segmentation was built. Requiring user input to determine both the image set's contrast state and initial seed points for segmentation growth, the algorithm functions independently thereafter, choosing methods and parameters for segmentation based on these data as well as each slice's axial depth. The algorithm checks its product by assessing the contour's size, connectedness, and similarity to a composite atlas created from the training kidney set, and adjusts methods and parameters based on the failures it encounters. Comparison of the autosegmentations to handsegmentations using the similarity metric indicated that the algorithm could perform acceptably well on new image sets when these data were within the scope of the algorithm's training. Use of the autosegmentation also compared favorably to handsegmentation when incorporated into an image guidance setting, producing surface-to-surface distances after registration similar to those obtained using the handsegmented surface. Further evaluation of the transformations obtained during registration showed low variability in the location of points after applying both the transform from the handsegmentation's registration and the transform from the autosegmentation's registration. These data support the use of the semi-automatic method as comparable to handsegmentation, and provide evidence that the semi-automatic method may be incorporated into an image guidance workflow.

REFERENCES

1. Ries LAG, M.D., Krapcho M, Stinchcomb DG, Howlader N, Horner MJ, Mariotto A, Miller BA, Feuer EJ, Altekruse SF, Lewis DR, Clegg L, Eisner MP, Reichman M, Edwards BK (eds). SEER Cancer Statistics Review, 1975-2005, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2005/, based on November 2007 SEER data submission, posted to the SEER web site, 2008.
2. Robson, C.J., B.M. Churchill, and W. Anderson, *The results of radical nephrectomy for renal cell carcinoma (Reprinted from J. Urol, vol 101, pg 297-301, 1969).* Journal of Urology, 2002. **167**(2): p. 873-875.
3. Belldegrun, A., et al., *Efficacy of nephron-sparing surgery for renal cell carcinoma: Analysis based on the new 1997 tumor-node-metastasis staging system.* Journal of Clinical Oncology, 1999. **17**(9): p. 2868-2875.
4. Lee, C.T., et al., *Surgical management of renal tumors 4 cm. or less in a contemporary cohort.* Journal of Urology, 2000. **163**(3): p. 730-736.
5. Huang, W.C., et al., *Chronic kidney disease after nephrectomy in patients with renal cortical tumours: A retrospective cohort study.* Journal of Endourology, 2007. **21**(2): p. 126-127.
6. Dunn, M.D., et al., *Laparoscopic versus open radical nephrectomy: A 9-year experience.* Journal of Urology, 2000. **164**(4): p. 1153-1159.
7. Lane, B.R., et al., *Comparison of laparoscopic and open partial nephrectomy for tumor in a solitary kidney.* Journal of Urology, 2008. **179**(3): p. 847-851.
8. Russo, P., *Open partial nephrectomy: an essential contemporary operation.* Nature Clinical Practice Urology, 2006. **3**(1): p. 2-3.
9. Galloway, R.L., *The process and development of image-guided procedures.* Annual Review of Biomedical Engineering, 2001. **3**: p. 83-108.
10. Friets, E.M., et al., *A Frameless Stereotaxic Operating Microscope for Neurosurgery.* IEEE Transactions on Biomedical Engineering, 1989. **36**(6): p. 608-617.
11. Galloway, R.L., R.J. Maciunas, and C.A. Edwards, *Interactive Image-Guided Neurosurgery.* IEEE Transactions on Biomedical Engineering, 1992. **39**(12): p. 1226-1231.
12. Roberts, D.W., et al., *A Frameless Stereotaxic Integration of Computerized Tomographic Imaging and the Operating Microscope.* Journal of Neurosurgery, 1986. **65**(4): p. 545-549.
13. Watanabe, E., et al., *3-Dimensional Digitizer (Neuronavigator) - New Equipment for Computed Tomography-Guided Stereotaxic Surgery.* Surgical Neurology, 1987. **27**(6): p. 543-547.
14. Cash, D.M., et al., *Concepts and preliminary data toward the realization of image-guided liver surgery.* Journal of Gastrointestinal Surgery, 2007. **11**(7): p. 844-859.
15. Herline, A.J., et al., *Image-guided surgery - Preliminary feasibility studies of frameless stereotactic liver surgery.* Archives of Surgery, 1999. **134**(6): p. 644-649.
16. Solomon, S.B., et al., *Real-time bronchoscope tip localization enables three-dimensional CT image guidance for transbronchial needle aspiration in swine.* Chest, 1998. **114**(5): p. 1405-1410.

17. Baumhauer, M., et al., *Soft tissue navigation for laparoscopic partial nephrectomy.* Int J CARS, 2008. **3**: p. 307-314.
18. Lunsford, L.D., *A Dedicated Ct System for the Stereotactic Operating-Room.* Applied Neurophysiology, 1982. **45**(4-5): p. 374-378.
19. Cash, D.M., et al., *Incorporation of a laser range scanner into image-guided liver surgery: Surface acquisition, registration, and tracking.* Medical Physics, 2003. **30**(7): p. 1671-1682.
20. Kim, J.K., et al., *Differentiation of subtypes of renal cell carcinoma on helical CT scans.* American Journal of Roentgenology, 2002. **178**(6): p. -.
21. Heidenreich, A. and V. Ravery, *Preoperative imaging in renal cell cancer.* World Journal of Urology, 2004. **22**(5): p. 307-315.
22. *The Insight Segmentation and Registration Toolkit*. [cited 2008; Available from: www.itk.org.
23. Pham, D.L., C.Y. Xu, and J.L. Prince, *Current methods in medical image segmentation.* Annual Review of Biomedical Engineering, 2000. **2**: p. 315-+.
24. Newman, T.S., et al. *A Volumetric Segmentation Technique for Diagnosis and Surgical Planning in Lower Torso CT Images*. in *Proceedings of ICPR*. 1996.
25. Osher, S. and J.A. Sethian, *Fronts Propagating with Curvature-Dependent Speed - Algorithms Based on Hamilton-Jacobi Formulations.* Journal of Computational Physics, 1988. **79**(1): p. 12-49.
26. 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, 2004. **8**(3): p. 217-231.
27. Adalsteinsson, D. and J.A. Sethian, *A Fast Level Set Method for Propagating Interfaces.* Journal of Computational Physics, 1995. **118**(2): p. 269-277.
28. Leventon, M.E., et al. *Level Set Based Segmentation with Intensity and Curvature Priors*. in *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*. 2000. Hilton Head SC.
29. Rousson, M. and N. Paragios, *Shape Priors for Level Set Representations*, in *Lecture Notes in Computer Science*, A. Heyden, Editor. 2002.
30. Osher, S. and R.P. Fedkiw, *Level set methods: An overview and some recent results.* Journal of Computational Physics, 2001. **169**(2): p. 463-502.
31. Viera, A.J. and J.M. Garrett, *Understanding interobserver agreement: the kappa statistic.* Family Medicine, 2005. **37**(5): p. 360-363.
32. Zijdenbos, A.P., et al., *Morphometric Analysis of White-Matter Lesions in Mr-Images - Method and Validation.* IEEE Transactions on Medical Imaging, 1994. **13**(4): p. 716-724.