Automatic Segmentation of the Human Abdomen on Clinically Acquired CT

By

Zhoubing Xu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May, 2016

Nashville, Tennessee

Approved:

Bennett A. Landman, Ph.D.

Benoit M. Dawant, Ph.D.

Benjamin K. Poulose, M.D.

Richard G. Abramson, M.D.

Jack H. Noble, Ph.D.

ACKNOWLEDGMENTS

I am deeply thankful for my Ph.D. career at Vanderbilt, especially as member of MASI lab. I am so lucky to meet all of those fantastic people. First, the lab leader, and my adviser, Bennett Landman – you have been giving me so much that I will be grateful forever. You brought me here when I almost lost my hope to get to Vanderbilt, you guide me how to be academically professional hand-by-hand, and most importantly, you make me believe in myself of being successful. To my lab mates, Andrew Asman – you are an incredible senior and a great exemplar for me to pursue; Xue Yang – without your help, I could never adapt myself to the lab in such a short time; Frederick Bryan, Benjamin Yvernault, Andrew Plassard, Rob Harrigan – there is always something to learn, and it is always fun to be around you; Yuankai Huo, Shikha Chaganti, Katrina Nelson, Shunxing Bao, and Sahil Panjwani – you make our lab the warmest atmosphere to work in; Steven Damen, and Charles Olivier – I can never appreciate too much for your technical support; Carolyn Lauzon, Kelsie Covington, Wade Allen, Bo Li, Swetasuda Panda, and Prasanna Parvathaneni – it is an honor to work hard together with you; Christopher Lee, Ryan Burke, Adam Gertz, and Jiaqi Liu – you are the greatest and most productive helpers, and you make me proud. For my other friends at Vanderbilt, I feel sense of belonging by sharing stories, experiences, and happiness with you. You make me regard Nashville as my second hometown.

I would like to delicate my achievements during Ph.D. to my family who always give me the unconditional love. To my dad and mom – I can never return your all-around support for doing this Ph.D. To my wife – you are the best gift for me, I cannot imagine making this far without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PART 1

INTRODUCTION

1. Background

The human abdomen is an essential, yet complex body space clinically. Bounded by diaphragm superiorly and pelvic inferiorly, supported by spinal vertebrae, and protected by muscular abdominal wall, the abdomen contains organs involved with blood reservation, detoxification, urination, endocrine, and digestion, and covers many important arteries and veins. Computational Tomography (CT) scans are routinely taken for the diagnosis and prognosis of abdomen-related diseases, such as the pathological injuries or changes of abdominal organs, and the abnormal extrusion in the abdominal wall, i.e., ventral hernia (VH), so that internal anatomic structures can be evaluated qualitatively without opening surgery.

Qualitative assessment could cause substantial subjective variations in treatment delivery for diseases. For example, to date, there is no standardized method for VH classification and repair that consistently and efficiently describes hernia characteristics. Failure rates of VH repair stay high, with recurrence rates estimated at between 24 and 43 percent [5]. Therefore, quantitative analysis becomes necessary, not only for precise measurements, but also for modeling a systematic degree of disease severity based on all measurable characteristics of specific structures, so that surgeons and radiologists can act accordingly.

Segmentation on CT images provides a computational representation of the structures of interest for accessing the characteristics (e.g., width, size, density), and thus establish a foundation for quantitative analysis. Manual delineation by anatomical experts is the most straightforward approach to extract the desired anatomic structure; however, it is impractical to have individual anatomical experts label large-scale clinical studies due to the time and costs. Further, manual segmentation inherently suffers from

intra- and inter- variability given the highly variable appearances of pathological structures in clinical CT [6]. Crowd-sourced, highly paralleled collaborative labeling has the potential to exploit the "wisdom of the crowd" (e.g., random volunteers with minimal training), and yield equivalent accuracy and efficiency to that of particularly "wise" individuals (e.g., surgeons, radiologists) with less costs and restrictions [7-9].

Fully automated segmentation on large-scale clinical data has been the target of intense efforts for decades. While extensive studies in brain segmentation have been developed [10-13], and targeted for therapeutic modifications and as surrogate endpoints in clinical trials [14, 15], abdominal structure segmentation has begun to draw more attention in recent years.

Abdominal organs have been the main focus. Probabilistic atlases are commonly built from co-registering the images with existing segmentations to characterize the spatial variations of the abdominal organs [16, 17]. Spatial standardization techniques have been used to regularize the inter-subject variations through landmark-based normalization, and yield patient-specific probabilistic atlases [18-20]. Statistical shape model approaches were developed to represent the anatomic shape changes of individual organs [21, 22] and / or the inter-organ hierarchies [23]. Graph-based methods provide an opportunity to combine shape and spatial models with other *a priori* information (e.g., appearance, contrast-enhancement), and yield global optimal segmentations [24-27] Multi-atlas approaches were newly adapted to abdominal organ segmentation [28]. Such techniques have been shown to enhance the generality and robustness of segmentations as opposed to that of the segmentations based on probabilistic atlases or statistical models that may be hampered by within-population variations [13].

Segmentation of the abdominal wall is also of interest in abdomen-related analyses. Zhang et al. aligned and transferred pre-defined muscular models at multiple key positions according to anatomic knowledge, and used active contour model methods to smooth the muscle segmentation [29]. Chung et al. trained a shape model in conjunction with an appearance prior from manually segmented images to regularize the registration from a binary mean muscle shape to the target image using PCA-encoded Free Form Deformation [30]. Ding et al. used a 3D flipping-free deformable model to register the inner

boundary of the wall for ease of segmentation and visualization of abdominal organs [31]. Yao et al. segmented the outer surface of the abdominal wall to separate subcutaneous and visceral adipose tissue by fuzzy c-means clustering and active contour models [32]. Zhu et al. provided an interactive approach to remove the entire the abdominal wall to reduce the sliding motion effect on the non-rigid registration of abdominal images [33].

While all efforts above show different degrees of success in extracting abdominal structures, there has been very limited dedication to a general-purpose solution for clinical use. In clinically acquired CT, additional variations from the imaging formats (e.g., field of view, resolution), contrast phases, and artifacts greatly challenge the robustness of automatic segmentation. Further, the pathological conditions (e.g., obesity, cancer, hernia, atrophied / enlarged / ablated / implanted organs) are commonly observed. The capability of handling the segmentation of large-scale clinical data with all these problems included defines a critical perspective of the clinical significances for the segmentation techniques.

2. Overview

In this dissertation, we develop algorithms for robust automatic segmentation of abdominal organs and abdominal wall in clinically acquired CT. The rest of the document is organized as follows. In **Chapter I**, the abdominal anatomy on CT imaging is described with the focus on the challenges of segmentation for clinical analysis. **Chapter II** reviews preliminary knowledge of image segmentation techniques. **Chapters III ~VI** focus on abdominal wall. **Chapter III** establishes a detailed, reproducible anatomical labeling protocol for ventral hernia characterization on CT scans with abdominal wall surfaces, fascia boundaries, and bony landmarks as a foundation for clinical endpoint prediction. **Chapter IV** develops an immersive virtual reality system to visualize abdominal wall and associated structures on abdominal CT of ventral hernia patients for better anatomical understanding. **Chapter V** integrates texture analysis into level set approach to improve the robustness of the segmentation of the outer surface of the abdominal wall given the commonly observed irregular structures induced by ventral hernias. **Chapter VI** presents an augmented active shape model framework to robustly extract the whole

abdominal wall and accurately measure the subcutaneous and visceral fat area. **Chapters VII ~XI** investigate abdominal organs. **Chapter VII** starts with a single organ (spleen), where implicit shape models are integrated into multi-atlas segmentation to constrain the outcome. Chapter VIII proposes a new atlas selection technique integrated with context learning, and establishes an automatic multi-atlas segmentation framework for multiple abdominal organs. **Chapter IX** designs a pipeline combines computer-assisted segmentation and manual outlier correction, and investigates its robustness and efficiency in deriving spleen volumes on a longitudinal clinical dataset. **Chapter X** evaluates six academically poplar registration methods on abdominal CT, provides recommendation for registration-based applications in abdomen, and suggests directions for future development of abdomen-specific registration tools. **Chapter XI** integrates an organ localization technique into the multi-atlas multi-organ segmentation framework, and demonstrates improved performances. **Chapter XII** concludes the main contributions of the dissertation, and discusses potential opportunities on future efforts.

3. Contributions

This dissertation aims to booster the clinical significance of structural abdominal segmentation in CT imaging. Thus, our focus covers not only (1) robust automatic segmentation, but also (2) effective clinical application based on the segmentation. The ultimate goal of this dissertation is to provide a general image processing pipeline to capture the structures of interest within the abdomen (i.e., abdominal organs and abdominal wall), and thus serve for clinical treatment with quantitative analysis and enhanced visualization. Specifically, the contributions of this dissertation are:

1. We proposed a framework to provide objective criteria for clinical ventral hernia analyses, where robust manual labeling protocol is designed to guide the automated processes in further study including the segmentation of abdominal wall, and the derivation of hernia-related characteristics.

2. We designed a system to render the segmentation of abdominal structures along with the original medical data in an interactive virtual reality environment, and thus offer the opportunities for improved understanding of the abdominal anatomy.
3. We presented an automatic segmentation for the outer surface of the abdominal wall on CTs of hernia patients. An edge-based level set method integrated with texture analysis was proposed to extract the anterior abdominal wall, which provided a more robust segmentation with less mean surface errors compared to other published methods.
4. We developed a robust automatic segmentation framework for the whole abdominal wall that extracts both the outer and inner surfaces and extends from the xiphoid process to the pubic symphysis. The framework used an augmented active shape model that combined multi-atlas label fusion and level set techniques, and enables the measurement of subcutaneous and visceral fat areas with close correlation to those derived from manual segmentation.
5. We built an implicit active shape model of spleen, and integrated into the multi-atlas label fusion framework to provide more accurate spleen segmentation in cases of problematic registrations, where parametric (active shape model) and non-parametric (multi-atlas label fusion) methods were seamlessly combined.
6. We generalized a theoretical framework of atlas selection to account for exogenous information (e.g., from separate models of tissue likelihood based context features), combined it with the state-of-the-art fusion techniques, and presented automatic multi-organ segmentation with significantly less dependence on non-robust registration.
7. We provided a robust and efficient pipeline for clinical use in spleen volume estimation that deployed automatic computer-assisted segmentation and manual outlier correction, and demonstrated its superiorities over traditional clinical approaches.
8. We evaluated six state-of-the-art registration methods based on their performances on 100 abdominal CT scans, and suggested potential directions for developing registration tools tailored for abdomen.

9. We investigated random forest techniques for localizing abdominal organs on CT scans, and proposed organ-wise multi-atlas labeling on the localized regions of interest that yielded better accuracy and less computational time over the traditional body-wise multi-atlas segmentation.

4. Previous Publications

Many contributions of this dissertation have been published. A framework is proposed to provide quantitative criteria based on a manual labeling protocol for ventral hernia analysis [34, 35]. An interface is designed to provide enhanced visualization and better understand of abdominal anatomy with a fully interactive virtual reality environment [36]. Automatic segmentation of the outer surface of abdominal wall is shown improved with texture analysis on data with excessive pathological conditions [37, 38]. Shape constraint [39] is shown to be critical for automatic segmentation of abdominal organs. Context learned atlas selection, image registration, and organ localization is demonstrated to improve the multi-atlas segmentation framework for abdominal organs [40-45].

CHAPTER I

ABDOMINAL ANATOMY ON CT IMAGING

1. Overview

The scope of the dissertation mainly covers the segmentation of abdominal organs and abdominal walls on CT scans. Thus understanding the anatomy in the abdomen on CT imaging is critical for that (1) we can tell what anatomical structures and which properties of structures are important to specific clinical applications; (2) we can tailor the techniques for the context features of the anatomical structures to segment. In this chapter, we review the abdominal anatomy on CT imaging with the focus on the challenges of segmentation for clinical analysis.

2. General Issues of CT Imaging

Computational Tomography (CT) is a routinely used scanning technique in clinical analysis using computer-processed X-rays to generate a series of tomographic image slices that allow the user to see the interior of human body for the specific scanned area without opening surgery. The CT images, often reformatted into a three-dimensional volume, demonstrate various anatomic structures with high-contrast resolution based on the differences in the physical density to absorb/block the X-ray beam between tissues, and the intensity of CT images, i.e., Hounsfield (HU) scale, has relatively consistent range for specific tissues (e.g., bones appears much brighter than fat tissues in CT scans).

Despite the inherent high contrast, CT scans for abdomen-related clinical analysis can vary from many perspectives, which make it challenging for robust automatic segmentation of abdominal structures on large-scale study.

2.1. Field of Views (FOVs)

Various FOVs exist among abdomen-related CT scans.

Along the cranial-caudal axis, some scans may range from the top of pelvis to the bottom of ribs (abdomen CT), while others may have larger extensions, i.e., from the top of sternum to the middle of femurs (thorax-abdomen-pelvis CT). A typical range of mostly taken abdominal CT scans covers the xiphoid process for its top, and the femur heads for its bottom (abdomen-pelvis CT), which effectively includes all abdominal viscera, the whole abdominal wall, a large portion of pelvis, spinal vertebrae, and rib cage, and the bottom end of lungs and heart. This diversity of the vertical coverage brings challenges of structure localization and spatial alignment between scans (e.g., registration).

Variations of FOV for in-plane slices are also observed. Sometimes, especially for those with obesity problems, the cross-sectional area of the patients' bodies is larger by the FOV of CT scanners so that a circle constraint can be seen on each axial slice. There are also cases that the patients are in contact with the scanner bore, which may cause streak artifacts on images (**Figure I.1a**).

In addition, the resolution of clinically acquired CT scans can also vary substantially. Among the cohorts that we have been studying with, the slice thickness (along the cranial-caudal axis) approximately ranges from 1.5 mm to 7.0 mm, while the in-plane resolution ranges from 0.5 x 0.5 mm to 1.0 x 1.0 mm. Conventionally, the voxels of the CT volumes are highly anisotropic (1.0 x 1.0 x 3.0 mm), the large slice thickness makes it difficult to yield smooth 3-D surfaces of anatomic structures.

2.2. Contrasts

Radiocontrast agents are often used in clinical CT to improve the visibility of internal anatomic structures in X-ray. Intravenous (IV) contrast is the mostly used medical contrast medium for abdomen CT scanning. It is an iodinated contrast agent given through a vein to help highlight abdominal organs, and detect tumors, infections, and vessel diseases. Sometimes, *per os* (PO) contrast, as a dilute iodinated contrast, is taken orally by patient to enhance the visibility of the gastrointestinal tract.

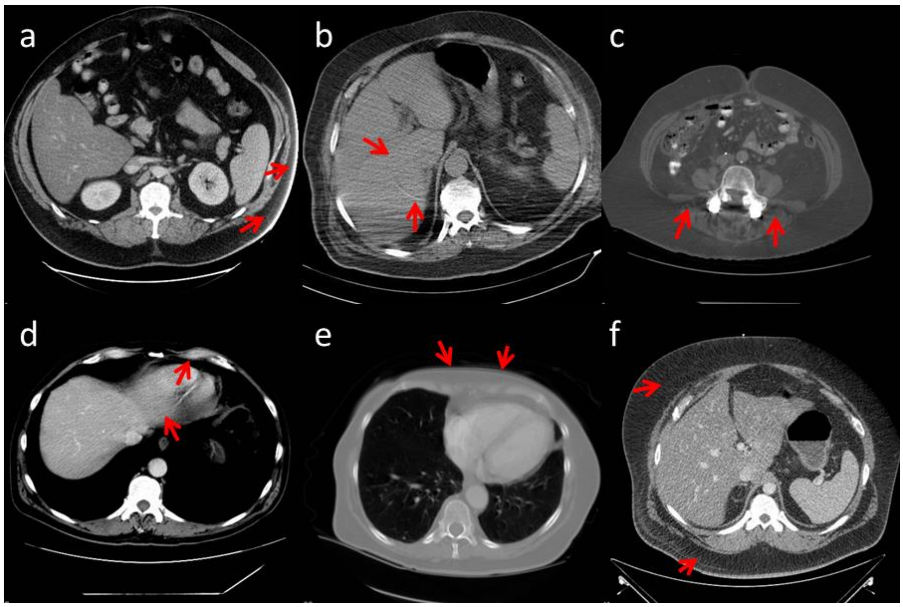

Figure I.1. Common CT artifacts: (a) streak artifact caused by body-scanner contact; (b) ring artifact; (c) windmill streak; (d) partial volume effect; (e) motion artifact; (f) streak artifact

The timing of taking a scan after the contrast is given is critical to the appearance of CT images. Early arterial phase, late arterial phase, and portal venous phase are the three typical IV contrast phases in terms of the propagation of the contrast agent through the vessel system, giving different structures highlighted as opposed to non-contrast phase where no contrast agent is given before scanning.

The distinct appearances (intensity distributions) of different contrast phases can undermine the robustness of intensity-based segmentation techniques. On the other hand, the relative contrast enhancement has been used as extra information to improve segmentation when scans of multiple contrasts are available [26, 27].

2.3. Artifacts

CT images are susceptible to a number of artifacts despite the generally faithful representations. Some typical artifacts (**Figure I.1**) are described below.

Streak artifacts are often seen around materials that block most X-rays. Bones, metal implants, female patients' bra wires can all contribute to the streak patterns in CT images.

Partial volume effects cause the blurring of the edges due to the scanner being unable to differentiate between a small amount of high-density material (e.g., bone) and a large amount of lower density (e.g., cartilage). They are most commonly seen when using highly anisotropic voxels, where the X-ray attenuation within each voxel is not as homogeneous as the reconstruction assumes.

Motion artifacts appear as blurring and/or streaking on image, caused by movement of patient being scanned; Ring artifacts are usually caused by mechanical detect fault or mis-calibration, shown as "rings" within image; Windmill streaking appearances are seen when the scanner detectors intersect the reconstruction plane.

3. Ventral Hernia and Abdominal Wall

A ventral hernia (VH) is not an intrinsic bodily structure, but an abnormal protrusion through a defect in the anterior abdominal muscular wall. Thus the abdominal wall is an important anatomic structure to characterize ventral hernias, where the muscular groups and/or fascial connections are typical extrusion spots of VHs, and several skeletal landmarks can be used to localize VHs on the abdominal wall. In **Chapter III**, we will propose a detailed labeling protocol based on the abdominal wall structures for hernia-oriented clinical analysis; here we first introduce the related anatomies (**Figure I.2**) briefly, and then describe several typical challenges for abdominal wall segmentation on clinically acquired CT.

3.1. Muscular Groups

The *rectus muscles* form the anterior-most muscular layer. The rectus consists of a pair of muscular columns oriented along the dorso-ventral axis. They are connected superiorly at the xiphoid process and extend inferiorly to the pubis.

There are three main muscles that make up the group of *oblique abdominal muscles* the external oblique, the internal oblique, and the transverse abdominal. These three muscles attach superiorly to the 5th through 12th ribs, inferiorly to the iliac crest, and extend across the anterior abdominal wall toward the rectus muscle along the midline. They form the lateral boundaries of the abdominal wall. Although the

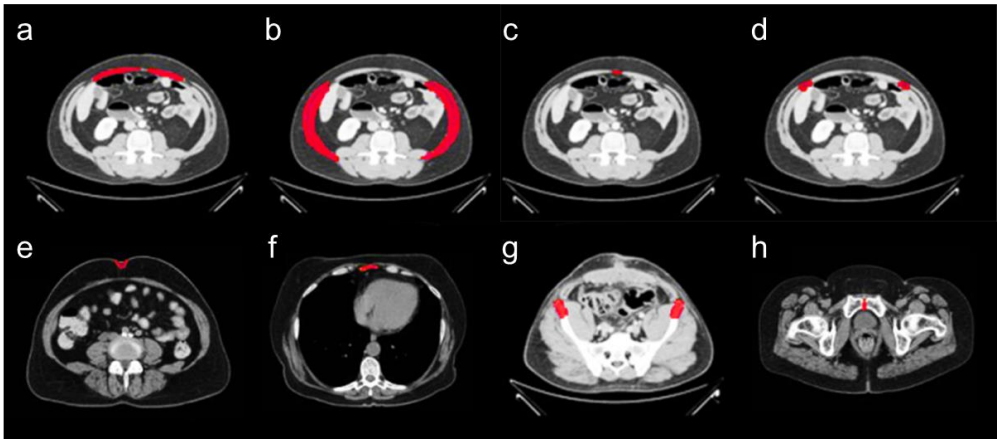

Figure I.2. Abdominal wall and related landmarks: (a) rectus muscles; (b) oblique abdominal muscles; (c) linea alba; (d) linea semilunaris; (e) umbilicus; (f) xiphoid process; (g) anterior superior iliac spines; and (h) pubic symphysis.

three lateral muscles can be well recognized, they are not consistently visually separable throughout their extent on clinical CT — especially at the terminations.

3.2. Fascial Connections

The *linea alba* runs between the two sections of the rectus muscle. It consists of almost exclusively connective tissue, including the fused aponeuroses of the rectus and the oblique muscular groups. It forms the anterior-most midline and extends from the xiphoid process superiorly to the pubis inferiorly. The linea alba is easily located on a CT image after identification of the rectus muscle.

The *linea semilunaris* runs directly aside the left and right boundaries of the rectus muscle, beginning at the termination of the oblique muscles. It consists of the aponeuroses of the oblique muscles. It begins superiorly at the 9th rib and extends inferiorly to nearly join at the pubic tubercle. On CT, the linea semilunaris can be identified by the anterior termina of the oblique abdominal muscles.

The *Umbilicus* is the small depression of the navel or "belly-button", located anteriorly along the linea alba slightly superior to the pelvis. On CT, the umbilicus can be seen as an indentation of the skin and subcutaneous tissue, and as a slight increase of intensity of the anterior dermal layer along the midline, anterior to the fascia.

3.3. Skeletal Landmarks

The *xiphoid process* is the inferior-most structure of the sternum. It is composed of ossified cartilage and is located medially, at the superior end of the linea alba. On CT, it can be found as a bright, bony structure directly in the midline, and its inferior tip forms the superior boundary of the anterior abdominal wall.

The left and right *anterior superior iliac spines* (ASIS) are the ridges along the superior curve of the pelvic bone. The anterior-most protrusion of the iliac curve is the anterior iliac spine. On CT, the anterior superior iliac spines are the anterior-most points of the bright, bony pelvis. They are useful as surface anatomical landmarks, which are palpable, as well as in defining the inferior-lateral boundary of the anterior abdominal wall.

The *pubic symphysis* is the cartilaginous joining of the two halves of the pubic bones, located medially and immediately posterior to the external genitalia. On CT, it can be found where the left and right bright, bony pubic structures nearly touch. It is useful in defining the inferior-most boundary of the abdominal wall.

3.4. Challenges of Abdominal Wall Segmentation on CT

We consider the segmentation of abdominal wall as to extract its inner and outer surfaces smoothly. To this end, we expect two general conditions for robust segmentation, (1) high contrasts for the boundaries of the abdominal walls, and (2) low variations for the surrounding tissues. However, these two conditions are not always met on clinical CT for hernia patients. For example, within the herniated region, the abdominal muscles (rectus / obliques) are highly stretched, and have inhomogeneous intensity, which can be hardly discriminated from the fat tissues around. Similarly, the fascial connections may have intensity similar to fat tissues, and cause the "leakage" problem for surface segmentation. On the other hand, lots of undesirable structures with similar intensity to abdominal muscles can appear around the abdominal wall, e.g., scar tissue, umbilicus, and sometimes skin tissue of skinny patients. The contrast-enhanced gastrointestinal tract and the speckle noise in fat increases the complexities from the

inside and outside of abdominal wall, respectively. In addition, the general CT artifacts discussed in section 2.3 also creates substantial segmentation difficulties, especially metal-induced streak artifacts that commonly observed from patients with metal implants.

4. Abdominal Organs

Underneath the abdominal wall, the human abdomen contains or partially contains a large portion of important organs, glands and vessels that tightly related to the functional quality of life. Generally on CT images, most of the abdominal viscera are not hard to identify, while the accurate segmentation of these viscera is challenging given the complicated environment in abdomen. As below, we briefly describe 11 groups of abdominal viscera on CT (**Figure I.3**), and discuss the abdominal viscera segmentation difficulties in general. Note that (1) we consider kidneys as abdominal organs, though strictly they are out of the abdominal cavity, (2) we skip the intestines because the serpentine tract, as a whole, is not commonly interested (as opposed to some pathological abnormalities, e.g., colon polyps), and too complicated to segment even by hand, and (3) heart and lung are beyond our scope here because they are in the thorax cavity, and usually not included entirely in abdominal CT scan.

4.1. Spleen

Spleen acts as a blood filter in human body primarily by removing old red blood cells and holding a reserve of blood. It is located in the left upper quadrant of the abdomen. On CT, the boundary of spleen is relatively easy to identify, except that sometimes it is in touch with the left kidney at the posterior end. Splenic artery and splenic vein diverge at the right hand side of spleen.

4.2. Kidney

The kidneys are essential organs in the urinary system, and serve the body as a natural filter of the blood, and remove wastes to the urinary bladder. They are located at the rear of the abdominal cavity in the retroperitoneum. On a regular CT scan in portal venous contrast phase, the intensity distribution of kidneys is slightly higher than, but share substantial overlap with that of the liver and spleen nearby.

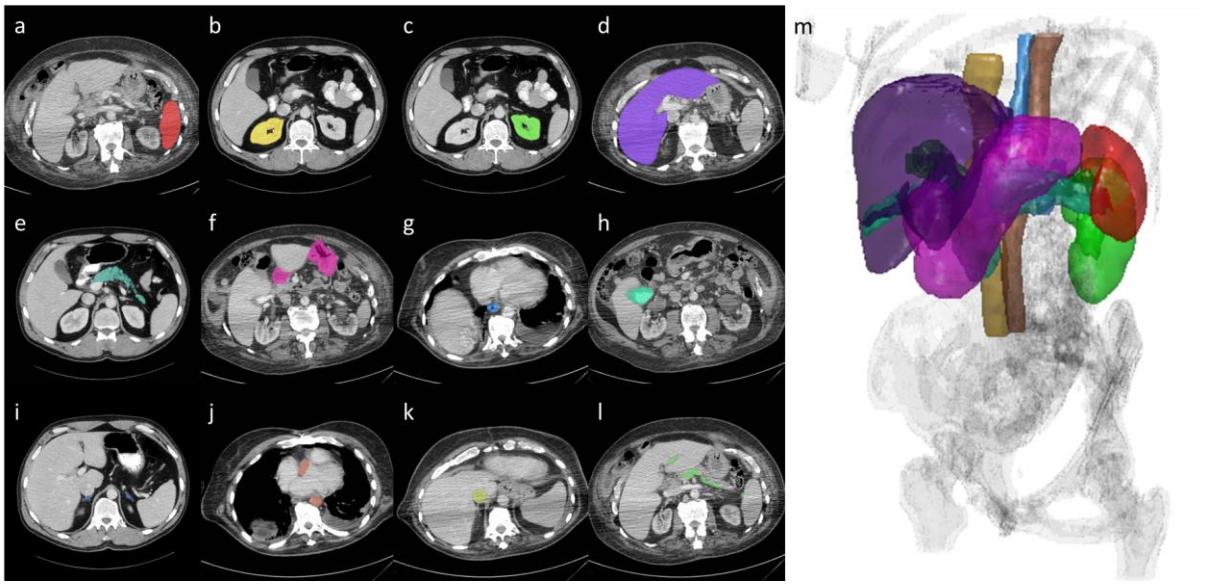

Figure I.3. Abdominal Viscera in CT: (a) spleen; (b) right kidney; (c) left kidney; (d) liver; (e) pancreas; (f) stomach; (g) esophagus; (h) gallbladder; (i) adrenal glands; (j) aorta; (k) inferior vena cava; (l) portal vein and splenic vein; (m) surface rendering of abdominal viscera.

Renal artery, renal vein, ureter, and renal pelvis loosely fill the concave surfaces in the middle of a kidney.

4.3. Liver

Liver has a wide range of functions, including detoxification, protein synthesis, and production of biochemicals necessary for digestion. It lies below the diaphragm in the abdominal-pelvic region of the abdomen. It is located in the right upper quadrant of the abdominal cavity. Some other anatomic structures are situated tightly with liver. Gallbladder is usually embedded underneath the middle of two liver lobes; inferior vena cava passes though the right-posterior end of the upper liver; portal vein extends branches into liver from its right hand side. A liver can often be observed in touch with the inner surface of the abdominal wall.

4.4. Pancreas

Pancreas is both an endocrine gland producing several important hormones, and a digestive organ, secreting pancreatic juice containing digestive enzymes that assist the absorption of nutrients and the digestion in the small intestine. The pancreas can be essentially divided into three parts: head, body, and tail. The head lies within the concavity of the duodenum. The body lies behind the stomach, and extends along the splenic vein. The tail is on the left end of the pancreas, lying towards the spleen. Pancreas has a fuzzy texture, and a slightly darker intensity than spleen and liver, its head part can be very confusing to distinguish from duodenum.

4.5. Stomach

Stomach is a muscular, hollow, dilated part of the digestion system which functions as an important organ of the digestive tract. The stomach is located between the esophagus and the small intestine. The gastrointestinal junction is divided at where the digestive tract looks pinched (pyloric sphincter). Stomach is on the left upper part of the abdominal cavity. The top of stomach lies against the diaphragm. A substantial part of stomach may appear as hollow, while the rest part is masticated food; both of them are enclosed by stomach wall. The shape of stomach can vary a lot in terms of food consumption and the existence of gastric distension.

4.6. Esophagus

Esophagus is an organ which consists of a muscular tube through which food passes from the pharynx to the stomach.

4.7. Gallbladder

Gallbladder is a small organ where bile is stored before released into the small intestine. The gallbladder is a hollow system that sits just beneath the liver. Normal gallbladder looks darker than liver on CT. Some patients may have gallstones, which appears as bright speckles in the gallbladder.

4.8. Adrenal glands

Adrenal glands are endocrine glands that sit at the top of the kidneys, chiefly responsible for releasing hormones in response to stress. They are wishbone-shaped structures with similar intensity and texture as pancreas, and situated bilaterally in the retroperitoneum superior to the kidneys. The right adrenal gland is usually close to the left hand side of liver, and left adrenal gland is be nearby the pancreas tail and spleen.

4.9. Aorta

Aorta distributes oxygenated blood to all parts of the body through the systemic circulation as the largest artery in the human body, originating from the left ventricle of the heart and extending down to the abdomen, where it bifurcates into two smaller arteries (the common iliac arteries). Aorta runs alongside of the vertebral column on its anterior side. Metal stents can be sometimes observed implanted in aorta, appearing as bright spots on the boundary of aorta.

4.10. Inferior Vena Cava (IVC)

IVC is the large vein that carries de-oxygenated blood from the lower half of the body into the right atrium of the heart. It is posterior to the abdominal cavity and runs alongside of the vertebral column on its right side (i.e. it is a retroperitoneal structure). It enters the right atrium at the lower right, back side of the heart. A substantial segment of IVC passes through liver, where its boundary can be barely observed.

4.11. Portal Vein and Splenic Vein

The hepatic portal vein is a blood vessel that conducts blood from the gastrointestinal tract and spleen to the liver. This blood is rich in nutrients that have been extracted from food, and the liver processes these nutrients. The hepatic portal vein is formed by the union of the superior mesenteric vein and the splenic vein. The portal vein and splenic vein together connect liver and spleen.

4.12. Challenges of Abdominal Viscera Segmentation

Generally, variations from multiple sources on clinically acquired abdominal CT images are the main obstacles of robust automatic segmentation of abdominal viscera. In other words, an approach works well on segmenting one subject might not be applicable to another one with much different appearance.

Firstly, the human abdomen is a highly deformable system. Even though constrained within a certain range, the spatial locations of most abdominal viscera can vary along with the fat deformation over different patients or different poses of the same patient. The stomach can also slide in a large extent by huge amount of food consumption before scanning and/or gastric distension. Some patients (usually suffer obesity) may have the viscera got in touch with each other firmly, which makes it even more challenging to segment each individual organ since most viscera share large overlap of intensity distribution on CT.

Next, the contrast phase of CT scanning also contributes to the inter-scan variations. As discussed in section 2.2, the appearances of the abdominal viscera are highly correlated with the contrast timing of the scan. While various phases are used clinically, the portal venous phase is the most favored IV contrast phase because it increases the contrast of some major abdominal organs, e.g., liver, spleen, and kidneys, as opposed to the non-contrast phase. If PO contrast is taken, the stomach and intestines become highlighted. This might be undesirable when segmenting other organs given the inherent large variations of these structures.

Lastly, abnormalities, or pathological conditions are more than commonly observed on clinical CT scans. The cancer or metastasis tissue shows darker than the healthy tissues of viscera; the degenerated and atrophied organs appear much smaller than usual; some patients might have the gallbladder ablated, while others might have an implanted artificial kidney. It is extremely difficult to adapt automatic segmentation methods to these very specific cases.

5. Other Related Anatomy

Some other anatomic structures can be supportive to the clinical analysis of abdominal wall and abdominal viscera.

The skin surface defines the boundary of human body. Combining with the abdominal wall surface, the volume of subcutaneous and visceral fat can be measured. These two metrics are very popularly used in recent clinical studies. The body mask enclosed by the skin surface can be acquired by intensity clustering [32].

The bony structures around the abdomen region, including ribs, pelvis and spinal vertebrae, are natural landmarks for localizing abdominal viscera. Especially, the segments of spinal vertebrae, nicely defines the human abdomen with multiple levels along the cranial-caudal axis [46, 47]. However, sophisticated automatic segmentation of bony structures will not be discussed in this dissertation.

CHAPTER II

SEGMENTATION PRELIMINARIES

1. Overview

Image segmentation is a process that partitions digital images into multiple segments, and usually used to extract the objects of interest on images. This process is equivalent to assign a distinct label to each set of pixels or voxels that represents a partitioned segment over the entire image (e.g., "0" for the background, and "1" for the foreground object of interest). The analyses on the segmented objects become easier with the simplified representation. Once segmented, the object of interest can be identified with enhanced visualization; its geometric properties (e.g., width, size) can be derived from the partitioned region; the constituent properties (e.g., density, texture) can be analyzed based on the image context under the partition. With these functionalities, medical image segmentation is getting more and more involved into daily clinical operations with evolutional significance (e.g., computer-aided diagnosis image-guided surgery, and post-operative treatment). In addition, robust automatic segmentation becomes very necessary for large-scale clinical study in place of the cost-consuming manual segmentation.

In this chapter, we review some preliminary knowledge of medical image segmentation. The review mainly covers the techniques used in the previous contributions described in the later chapters. Note that during the discussion below, CT is considered the default image modality, while most of the techniques are also applicable to other image modalities.

2. Image Features

Technically, the key of image segmentation is to generalize image features appropriately so that the image can be divided into meaningful partitions. For example, imagine that we are trying to extract an apple from a piece of white cloth beneath it in a picture, we might consider the color (intensity)

differences between apple and cloth, and the generic shape of an apple as the image features for segmentation. Having a good understanding of image features is critical to image segmentation. It helps us to answer the questions like (1) which image features provide the primary discriminant for specific segmentation tasks; (2) how multiple image features can be weighted and adapted for images with large variations; (3) when some image features should be included for refining the segmentation, though they cannot handle the segmentation by themselves. We consider that image features can be categorized into three classes discussed below.

2.1. Intensity-Driven Features

Most image segmentation methods operate based on intensity-driven features, including the image intensity, and other pixel-wise values derived from intensity.

2.1.1. Intensity

Intensity, or grayscale, is the most basic image feature, it is the value assigned to each pixel on image. If using linear intensity-darkness lookup table, intensity is basically what we see on images. Thus naturally, if the object to segment looks darker or brighter than the rest part of image, it could be identified by intensity, e.g., an intensity threshold would effectively extract the object from the background. Note that color digital images are made of multiple-channel grayscale images, but beyond our scope of discussion. Here we focus on single-channel images as most medical image modalities.

2.1.2. Edge

Most objects of interest are confined by edges (boundaries) in image. Thus identifying the edges of objects can help to extract the object. Sometimes, these boundaries are the actual target interests. The edges on images are mostly characterized by image gradient, while the gradient is a vector of the first derivatives of intensity along all image dimensions, or more commonly considered as the L^2-norm of the vector. Generally, strong edges have large gradient, blurry boundaries corresponds to moderate gradient,

while gradient over flat regions approaches zero. Lot of edge-based segmentation has been employing the discriminant effectivity of local gradient [48-52].

2.1.3. Texture

Texture can be considered as the appearance represented with repeated local patterns. Textures with different local patterns are distinguishable, e.g., we can tell the wood texture from that of the fabric. Texture analysis has been a long studied technique in computer vision [53], and has been applied to medical imaging [54]. Various approaches characterize textures differently, among which methods using co-occurrence matrices and Gabor filters are most commonly used. Co-occurrence matrix considers textures as the spatial distribution of intensities [55, 56]. It represents the second-order statistics of the image, where relationships between groups of two voxels are collected. Based on the co-occurrence matrix, Haralick et al. [57] proposed 13 features to describe local textures. On the other hand, Gabor filters extract features in frequency domain by passing images through multi-channel filter operators. With a bank of Gabor filters, frequencies and orientations at multiple scales can be collected from images as texture features for classification and segmentation [58-62]. The perceived local textures can be used as an alternative of / in addition to the intensity values for image segmentation [63-65].

2.2. Atlas-Based Features

Image atlas represents an existing segmented object (i.e., atlas label) along with the context image (i.e., atlas image), which provides an example for how the desired object should be segmented. Thus in medical imaging, atlas-based features are often used since the human anatomy is generalizable in terms of spatial distraction, shape and appearance.

2.2.1. Spatial Distribution

The principal information provided by image atlas is the spatial distribution of the structure of interest in the image context. Assuming two abdominal CT images are roughly aligned, and the liver on one of the image has been labeled, then we can infer that the liver should be at about the same place on

the other image. Spatial information is extremely helpful in a complicated environment where the structure of interest share similar intensity-based features with other undesired structures. Note the alignment of images is a necessary step in common tasks of atlas-based segmentation, and referred as image registration, which is discussed in section 3.4.

2.2.2. Shape and Appearance

The shape and appearance information can also be generalized from multiple image atlases. Still taking liver as an example, given various liver segmentations on multiple images, it is reasonable to assume that the liver shape on a new image is most likely within the variations of the existing liver segmentation (if there is no abnormality presented on the new image). This idea induces the theory of Active Shape Model (ASM, [66]), through which the modes of the variations of the object shape can be represented as a vector of shape parameters in an eigenspace constructed by Principal Component Analysis (PCA, [67]), and thus fitting a new shape corresponds to finding an optimal set of shape parameters (e.g., by propagating the shape surface to the boundaries [66]). By incorporating the underlying context, Active Appearance Model (AAM, [68]) characterizes the modes of appearances similarly, which provides additional references for segmentation.

2.3. Exogenous Information

Some exogenous information associated with images can also support the segmentation.

Patients' demography and medical history provide natural basis to classify their medical images. For example, female patients have broader pelvis than male ones, older people are more likely to have atrophy issues, and patients having opening surgery history may share similar patterns of scar tissue. If available, these types of information can reasonably divide the whole cohort into subsets with smaller variations, and hence provide improved segmentation tailored for each group.

In addition, some properties of images, e.g., contrast phase of CT scans, indicate the modes of image appearances. Learning the contrast enhancement between non-contrast and portal venous contrast phases as additional independent features was shown helpful to drive the image segmentation [26, 27].

3. Pre-Processing

In practice, image regularization is often required before segmentation due to (1) various field-of-views (FOVs), (2) inhomogeneous intensities, (3) excessive noise, and (4) non-alignment between target and atlas; otherwise the segmentation might hardly work. Thus some pre-processing techniques are applied to images to unleash these problems, and help the segmentation methods more applicable to images.

3.1. Region of Interest (ROI) Selection

When the desired object to segment is much smaller than the FOV of the entire image, an ROI selection might be necessary so that (1) redundant regions are removed, (2) confusing undesired structures are reduced, and (3) computational time becomes less.

Manual selection of ROI is an acceptable option as long as the structure of interest is easy to identify. Several approaches can be considered to automate the process. When image atlases are available, and the registrations are robust, an ROI can be considered as the union of the registered the atlas labels [69]. When some landmarks are easy to extract automatically, and they roughly covers the range of the structure of interest, the ROI in the form of bounding box can be derived from the coordinates of the landmarks. Recently, sophisticated localization techniques are presented [70, 71], where ROIs are robustly extracted by learning the image context.

3.2. Intensity Normalization

Intensity normalization is a regularization process on the range of intensity values. It is often used to enhance the contrast over a certain intensity range so that some objects become easier to identify on the image. Consider the original image I_0 with intensity range (a_0, b_0), the normalized image $I_1$with intensity range (a_1, b_1), then a commonly used linear normalization can be described as follows

$$I_1 = \frac{(I_0 - a_0)(b_1 - a_1)}{b_0 - a_0} + a_1 \tag{2.1}$$

In CT scans, linear intensity normalization might be used for the following situations with specific configurations of the desired intensity range after normalization: (1) extremely high intensity values are observed over the metal implants and some bony structures, which unnecessarily broadens the intensity range of image; (2) the bony structures need to be enhanced as a skeleton reference; and (3) the segmentation requires high contrast between muscle and fat tissue.

3.3. Smoothing and De-noising

Image noise can be considered as random local intensity variations in digital images. When image noise is present, the efficacy of image segmentation techniques, especially edge-based methods, can be greatly affected. Thus image smoothing is often necessary to remove the noise. In image processing, noise is considered as high frequency content in the image, thus low pass filters are used for smoothing.

Gaussian filters are mostly used because of its simple form and the fact that most image noise is assumed Gaussian distributed. In 1-D, the Gaussian filter has the form of

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{x^2}{2\sigma^2}) \tag{2.2}$$

where σ represents the standard deviation of the Gaussian filter. In practice, image smoothing is usually operated by the convolution between the original image and the filtering kernel. An extra variable is required to define the kernel size, as the digital approximation of the analogical filter. In addition, isotropic Gaussian smoothing for 2-D and 3-D images can be performed by convolving 1-D Gaussians for all dimensions one after another. When using Gaussian filters, some important image features (e.g., edges and other details) can be blurred at the same time as the noise is reduced. To address this problem, anisotropic diffusion [72] and a newly developed technique - sheetness filter [73] can be used to reduce noise without removing significant parts of images.

3.4. Image Registration

Image registration is a process to transform different images into the same coordinate space, or more naturally speaking, it is a transformation process that aligns the contents on different images. This

process has been increasingly used in healthcare and medical research [74, 75]. With registration, information associated with the registered images becomes the additional reference for the target image, which establishes the foundation for many modern image-processing techniques. Thus image registration can be considered as a pre-processing step for image segmentation, while the quality of registration is critical for atlas-based segmentation methods [76-78].

Theoretically, a registration is generally an iterative process that involves a target image (or fixed image) considered as the reference, a source image (or moving image) that is expected to align with the target image, a similarity metric (e.g., fiducial registration error (FRE, [79]), sum of squared differences (SSD), normalized cross-correlation (NCC, [80]), mutual information (MI, [81, 82]), normalized mutual information (NMI, [81, 82])) to measure the alignment degree between the two images, and a optimizer (e.g., gradient descent [83], Powell's method [84], Nelder-Mead [85]) to help the registration reach an optimum of the defined similarity metric efficiently along the iterative search over all degrees of freedoms (DoFs) of the image transformation (e.g., translation, rotation, scaling, shearing, deformation). An interpolator (e.g., nearest neighbour, linear [86], spline [87]) is needed to measure and assign intensity values on the image grids since the registered image voxels are usually located between grids after transformation.

In terms of the registration basis, image registration can be categorized into feature-based and intensity-based. Feature-based registration seeks to align the correspondent features (e.g., points, lines, surfaces) in images, through which determines the geometrical transformation, and then map the source and target images together [88, 89]. Intensity-based methods drive the registration by comparing the similarities of the intensity-based metrics [90, 91]. With regard to the transformation mode, image registration can be roughly classified as rigid, affine, and non-rigid, where different DoFs are allowed. Rigid registration permits only translation and rotation; affine registration [92] allows scaling and shearing in addition, while the transformation stays linear; non-rigid registration enables non-linear transformation [93], i.e., local warping, to align images, where the local warping can be modeled with thin-plate spline [94], free form deformation [95], diffeomorphisms [96], and etc.

4. Modern Segmentation Techniques

Several modern segmentation techniques mentioned in the later chapters are introduced below. We note that this is not a comprehensive review of image segmentation techniques; thresholding [97, 98], watershed [99], active contour models [48, 49], and graph-based segmentation [100, 101] methods are not currently covered.

4.1. Bayesian Methods

In the statement of Bayes' theorem,

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{2.3}$$

A is considered as an outcome event, B is an observed event, $P(A)$ is the prior probability (initial belief) that event A occurs, $P(B|A)$ represents a generative model that characterizes the probability distribution of the random observations given a latent outcome, $P(B)$ is the total probability of observation, and act as a partition function (normalization term), $P(A|B)$ represents the posterior probability of the outcome given observations.

The Bayes' theorem provides a very basic framework for image segmentation. Specifically, considering A as the segmentation decision, B as the image features, then the Bayes' theorem naturally determines the segmentation estimate based on the image features, which can be just a scalar of intensity value, or a feature vector that includes various image contexts. While in Eq. 3, the prior is usually estimated roughly and the partition function serves for normalization, the generative model is very critical for the posterior estimate. Typically, the generative model can be learned with examples, i.e., data with both observations and outcome, and then applied to Eq. 3 for the outcome estimate on newly observed data. This training-testing scheme with Bayes' theorem has been the basis of many supervised machine learning [102] and image segmentation techniques [103, 104].

In practical image segmentation, the generative model can be characterized by Gaussian mixture models (GMMs) [105, 106]. On a voxel-wise basis, let $\boldsymbol{x} \in \mathbb{R}^{d\times 1}$represent a d dimensional image feature

vector, $m \in \boldsymbol{M}$ indicate the segmentation estimate, where $\boldsymbol{M} = \{1, \dots, M\}$ is the set of possible outcomes. The probability of the observed features given the segmentation is m can be represented with the mixture of N_G Gaussian distributions,

$$f(\boldsymbol{x}|m = t) = \sum_{k=1}^{N_G} \frac{\alpha_{kt}}{(2\pi)^{\frac{d}{2}}|\boldsymbol{C}_{kt}|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_{kt})^T \boldsymbol{C}_{kt}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{kt})\right] \tag{2.4}$$

where $\alpha_{kt} \in \mathbb{R}^{1\times1}$, $\boldsymbol{\mu}_{kt} \in \mathbb{R}^{d\times1}$, and $\boldsymbol{C}_{kt} \in \mathbb{R}^{d\times d}$ are the unknown mixture probability, mean, and covariance matrix to estimate for each Gaussian mixture component k of each tissue type t by the Expectation-Maximization (EM) algorithm following [107]. After the generative model is learned from datasets with known segmentations, the segmentation estimate on unknown datasets can be by Bayesian expansion with extracted feature vectors.

$$f(m = t|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|m = t)f(m = t)}{\sum_{t'} f(\boldsymbol{x}|m = t')f(m = t')} \tag{2.5}$$

4.2. Clustering

Data clustering is process to partition observations into different clusters with minimum within-cluster sum of squares (WCSS) based on the observed features. It can be considered as an unsupervised counterpart of Bayesian methods, where no model training is required, instead a number of desired clusters should be defined. *k*-means clustering [108] is a basic clustering method, where after partitioning, each data point belongs completely to the assigned cluster. Fuzzy c-means (FCM) clustering allows data points with partial membership (fuzziness) of each cluster [109].

Consider $\mathbf{X}=\{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$ as a set of feature observations for an image with n voxels, where $\boldsymbol{x_j} \in \mathbb{R}^{d\times1}$represent a d dimensional image feature vector, fuzzy c-means clustering aims to divide the n observations into c sets with the cluster centers defined as $\boldsymbol{C} = \{\boldsymbol{c_1}, \boldsymbol{c_2}, \dots, \boldsymbol{c_c}\}$. Partition coefficients are measured as $\mu_{ij} \in [0, 1]$ for the cluster membership j of the observation i, where $i = 1, \dots, n, j = 1, \dots, c$. The goal is the minimize a objective function,

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{C} \mu_{ij}^{m} \left\| \boldsymbol{x_i} - \boldsymbol{c_j} \right\|^2, 1 \le m < \infty \tag{2.6}$$

where m is a real number no less than 1, defined as a fuzziness coefficient. The fuzzy partition is carried out through an iterative optimization with the update of the voxel-wise membership μ_{ij} and the cluster centers c_j

$$\mu_{ij} = \left(\sum_{k=1}^{C} \left(\frac{\left\| \boldsymbol{x_i} - \boldsymbol{c_j} \right\|}{\left\| \boldsymbol{x_i} - \boldsymbol{c_k} \right\|} \right)^{\frac{2}{m-1}} \right)^{-1} \tag{2.7}$$

$$c_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{m} \cdot \boldsymbol{x_i}}{\sum_{i=1}^{n} \mu_{ij}^{m}} \tag{2.8}$$

and the iteration with stop when the maximum update increment of μ_{ij} is less than a small number $\varepsilon \in (0, 1)$. Note that when $m = 1$, the membership u_{ij} converges to 0 or 1, and the FCM clustering approaches *k*-means clustering.

4.3. Active Shape Model (ASM)

ASM represents a parametric deformable model that statistically characterizes the shape variations from the training set. ASM is often built to regularize the shape of the desired structure while segmenting it [66].

Classic ASM employs point distribution model (PDM). The building of PDM requires a collection of corresponding points along the shape boundaries or salient features on the training sets, where the shapes from all datasets are usually aligned by registration to a normalized reference coordinate. The shape variations are modeled following a Principle Component Analysis (PCA) procedure. In a 3-D case, the coordinates of n shape points along the i^{th} of m training sets in the aligned coordinate are first defined as

$$\boldsymbol{x_i} = (x_{i1}, y_{i1}, z_{i1}, \dots, x_{in}, y_{in}, z_{in})^T \tag{2.9}$$

Then the mean shape can be given as

$$\bar{\boldsymbol{x}} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{x_i} \tag{2.10}$$

The modes of variations are given by the singular value decomposition (SVD) of the covariance matrix

$$\boldsymbol{U\Sigma U^T} = \frac{1}{\mathrm{m}}\sum_{i=1}^{m} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^T \tag{2.11}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with its diagonal values as the eigenvalues λ_k, and $\boldsymbol{U}$ denotes an eigenspace with each column as an eigenvector, i.e., one mode of variation $\boldsymbol{u_k}$, associated with the corresponding eigenvalue λ_k.The value of the eigenvalue indicates the dominance of its associated mode of variation, while the modes with relatively small eigenvalues are usually ignored due to their limited variances provided. *K* eigenvectors representing the most variances remain and form a eigenspace $\boldsymbol{U_K}$ with reduced dimensions. With the shape model built, a shape instance can be reconstructed by deforming the mean shape by a linear combination of eigenvectors

$$\boldsymbol{x} \equiv \boldsymbol{x}(\boldsymbol{b}) = \bar{\boldsymbol{x}} + \sum_{k=1}^{K} b_k \boldsymbol{u_k} \tag{2.12}$$

where b_k is denotes the shape parameter associated with its mode of variation.

For practical segmentation, ASM usually works by alternating over local profile fitting and shape regularizing. Specially, after initialization, a shape is first generated by searching around each point for a position that matches the modeled local profile better [22, 110], usually at strong edges [72]. Then the newly generated shape $\hat{\boldsymbol{x}}$ is constrained by the shape model

$$\hat{\boldsymbol{b}} = \boldsymbol{U_K}(\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}) \tag{2.13}$$

where $\hat{\boldsymbol{b}}$ represents the projected shape parameter that can generate a conformed shape by the shape model through Eq. 12. In addition, $\hat{\boldsymbol{b}}$ is often constrained within certain range of the variations of the shape model represented by the eigenvalues $\boldsymbol{\lambda}$.

Building training sets for ASM is very challenging and time-consuming, especially to provide corresponding points. Several techniques have been developed to generate corresponding points automatically [111, 112]. Some studies use signed distance function (SDF) over the whole image instead

of PDM to build ASM after all training sets are co-registered so that corresponding points are not required [113-115].

4.4. Level Set

Level set segmentation provides a numerical approach to track the surfaces of objects [116-118]. Level set methods is very similar to active contour models (ACM), e.g., snake [49], for that both methods define cost / energy functions to minimize as the criteria for segmentation, and derive Euler-Lagrange equation to numerically propagate the segmentation surface to approach to the objects of interest. Level set, however, has two main advantages over ACM: (1) it is free from parameterization of the propagating surface, and (2) it is easier to change topology to fit the underlying object (e.g., split and merge shapes). Level set segmentation is broadly used to extract smooth deformable surfaces of desired structures in images.

In the level set approach, an evolving surface is embedded as the zero level set of a higher dimensional level set function $\phi(x,t)$, and propagates implicitly through the temporal evolution of ϕ in terms of a given speed function F. $\phi(x,t)$ is defined as signed distance function (SDF) to the evolving surface with negative value inside the surface and positive outside. The speed function, which can be spatially varying, is usually determined by advection forces (e.g., constant inward or outward motion), intrinsic geometry (e.g., mean curvature), and image attributes (e.g., intensity, and its gradient). The temporal evolution of the level set function is usually described in the following form

$$\phi_t - F|\nabla\phi| = 0 \tag{2.14}$$

where $|\nabla\Phi|$ represents the normalized gradient of the level set function. Note that Φ_t is the Euler-Lagrange equation to update the surface.

Generally, level set segmentation is classified into two categories: region-based and edge-based.

Region-based level set methods rely on the global homogeneity of spatial localized features and properties [119-121], among which Chan-Vese algorithm [119] is one of the mostly used. Chan-Vese algorithm aims to propagate the curve by minimizing the variances inside and outside the curve while

constrained the total curve length. Consider $u_0(x,y)$ as a 2-D input image, $H(\cdot)$ as a Heaviside function, $\delta(\cdot)$ as a Dirac function, the derived Euler-Lagrange equation of Chan-Vese method can be written as

$$\phi_t = \delta(\phi)[v\,\kappa - (u_0 - c_1)^2 - (u_0 - c_2)^2] \tag{2.15}$$

where v is a parameter to constrain the length of the contour, $c_1$and c_2 are the average of u_0 outside and inside the current zero level set, respectively, and κ represents the curvature of ϕ that enforcing the smoothness of the curve.

$$c_1(\phi) = \frac{\int_\Omega\ u_0(x,y)\,H(\phi(t,x,y)dxdy}{\int_\Omega\ H(\phi(t,x,y)dxdy} \tag{2.16}$$

$$c_2(\phi) = \frac{\int_\Omega\ u_0(x,y)\,\big(1 - H(\phi(t,x,y)\big)dxdy}{\int_\Omega\ \big(1 - H(\phi(t,x,y)\big)dxdy} \tag{2.17}$$

$$\kappa = div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) \tag{2.18}$$

where Ω represents the whole domain of image. Note that $H(\cdot)$ and $\delta(\cdot)$ requires a continuous approximation in implementation to enable the actual curve evolution.

Edge-based level set methods drive the surface propagation through the local differences [50-52], typically the edges, in the image, among which Geodesic Active Contours (GAC) model [50] is very commonly used. Basically, GAC pushes the curve by an inward / outward normal pressure force, i.e., balloon force, and stops it based on an edge map defined by the image gradient.

$$\phi_t = g|\nabla\phi|(c + \kappa) + \nabla\Phi\cdot\nabla g \tag{2.19}$$

where κ is the curvature as in Eq. 18, c is a parameter that controls the strength of balloon force, g is called edge stopping function, proposed by Malladi et al. [122] to stop the level set evolution at the boundaries of objects, which can be typically measured for an image I in the form as follows

$$g = \frac{1}{1 + |\nabla \mathrm{G}_\sigma * I|^p} \tag{2.20}$$

where p controls the steepness of the edge stopping function, typically 1 or 2, and $G_\sigma$denotes a Gaussian smoothing (convolution) kernel with a standard deviation of σ. Back to Eq. 19, the balloon force is seen reasonably constrained by the edge stopping function as it can be diminished at strong edges. In addition, the term $\nabla\Phi\cdot\nabla g$ performs to further preserve the curve from propagating through the edges.

In addition, re-initialization of level set is usually required during the surface evolution to keep the pattern of SDF, and thus preserve the properties of the ϕ-related terms []. The standard re-initialization is solved by

$$\phi_t = sign(\phi_0)(1 - |\nabla\phi|) \tag{2.21}$$

where ϕ_0 denotes the SDF to be re-initialized, $sign(\cdot)$ indicates a sign function. Recently, Li et al. proposed an internal regularization energy term, i.e., $F_{reg} = \nabla^2\Phi - \kappa$, that enables the level set evolution free from additional re-initialization [123].

4.5. Multi-Atlas Label Fusion

As registration does not necessarily (is actually challenging to) align the atlas image perfectly at the region of interest, single-atlas-based segmentation is not robust, and almost solely depends on the quality of registration. Several studies have demonstrated that the combined information of multiple atlases registered to a single target image yields significantly improved segmentation [13, 124]. Multi-atlas label fusion is such a technique for transferring and combining information from multiple canonical atlases to target images via registration.

By definition, a natural workflow of multi-atlas label fusion includes multiple pair-wise registrations from atlas to target based on the intensity images, the propagation of the atlas labels (existing segmentations) to the space of the target image via the geometrical transformation that align the intensity images together, a process that fuses the propagated labels into one segmentation estimate, called label fusion. In practice, especially when the registrations are not robust due to large variations of the datasets, an additional process, i.e., atlas selection can substantially improve the segmentation by choosing the most appropriately registered atlases for fusion. Thus we focus on the three critical components in a multi-atlas label fusion framework: (1) registration (including intensity-based registration, and label propagation), (2) atlas selection, and (3) label fusion.

4.5.1. Registration

The image registration in multi-atlas label fusion aims to align the structures in the atlas images to those in the target image so that the atlas labels become valuable references to the segmentation estimate. By default, intensity-based similarity metrics including are used to drive the registration, where NCC, MI, NMI are qualified with the alignment involving intensity inconsistencies. To avoid getting to local minima, a coarse-to-fine scheme is often used, in which registrations are carried out between the images with downsampling levels, and the registered transformation between the coarser images are used as the initialization for further registrations between finer images. Similarly, rigid and affine registrations are often used to align atlas and target image roughly, followed by a non-rigid registration that allows local deformation to match up the details [28].

4.5.2. Atlas Selection

Atlas selection is a process to alleviate registration problems. Although in some sophisticated label fusion techniques, the global and local weighting of atlases can be considered as a fuzzy version atlas selection, the presence of the bad atlases can bias the evaluation of other atlases, and thus influence the fusion results [125]. Therefore, a crisp determination of whether an atlas should be included for label fusion becomes necessary, especially when substantial problematic registrations exist. Traditional image similarity metrics between the registered atlas images and the target image have been considered as the proxies of for atlas-target similarity, while the summations of these voxel-wise metrics over regions of interest are used as the selecting criteria [76, 77, 126]. On the other hand, Selective and Iterative Method for Performance Level Estimation (SIMPLE) [127]provides a compact solution for selecting atlases and fusing segmentation iteratively based on the overlapping regions of atlas labels with the intermediate fusion result, which demonstrates better correlation with the registration performances than the image similarity metrics [77].

4.5.3. Label Fusion

Label fusion attempts to estimate the underlying segmentation through combining individual estimates from multiple sources [128, 129]. It was proven that a collection of estimates only slightly better than chance could be fused (or "boosted") to form a single estimate with arbitrarily high accuracy [130-132]. In the context of multi-atlas label fusion, label fusion yields an estimated segmentation the target image and the multiple registered atlases (including both intensities and labels) [13, 124]. The simplest way to fuse the register atlases is through a majority vote of the registered atlas labels [129], where the mode label is selected at every voxel as the estimated segmentation for target. However, majority vote fails to assess both the likelihood of the estimate and the quality of the observed segmentations. Today, label fusion is an extremely popular topic for ongoing research.

In general, consider a target intensity image represented as a vector, $\boldsymbol{I} \in \mathbb{R}^{N\times 1}$. Let $\boldsymbol{T} \in \boldsymbol{L}^{N\times 1}$ be the latent representation of the true target segmentation, where $\boldsymbol{L} = \{0, \dots, L-1\}$ is the set of possible labels that can be assigned to a given voxel. Additionally, consider a collection of R registered atlases with associated intensity values, $\boldsymbol{A} \in \mathbb{R}^{N\times R}$, and label decisions, $\boldsymbol{D} \in \boldsymbol{L}^{N\times R}$. Throughout, the index variables i will be used to iterate over the voxels, s and l over the labels, and j over the registered atlases. Given this information, the goal of any label fusion framework is to accurately estimate the following probability density function:

$$W_{si} \equiv f(T_i = s|\boldsymbol{I}, \boldsymbol{A}, \boldsymbol{D}) \tag{2.22}$$

where W_{si} can be interpreted as the probability that the true label at voxel i is equal to label s given the provided contextual information. Then using a Bayesian expansion, Eq. 23 can be re-written as:

$$W_{si} = \frac{f(T_i = s) f(\boldsymbol{D}, \boldsymbol{A}|T_i = s, I)}{\sum_l f(T_i = l) f(\boldsymbol{D}, \boldsymbol{A}|T_i = l, I)} \tag{2.23}$$

where, $f(T_i = s)$ denotes the *a priori* distribution governing the underlying segmentation and $f(\boldsymbol{D}, \boldsymbol{A}|T_i = s, I)$ is the generative model that represents distribution governing the relationships between the observed atlas information and the latent target segmentation. Lastly, with the assumption that the observed atlas labels and the observed atlas intensities are conditionally independent [133]:

$$W_{si} = \frac{f(T_i = s) f(\boldsymbol{D}|T_i = s) f(\boldsymbol{A}|I)}{\sum_l f(T_i = l) f(\boldsymbol{D}|T_i = l) f(\boldsymbol{A}|I)} \tag{2.24}$$

where the information gained by direct incorporation of the atlas-target intensity relationships, i.e., $f(\boldsymbol{A}|I)$, is assumed to accurately approximates the complex relationships through the assumed conditional independence.

With Eq. 25, any label fusion method can be considered as the specification of (1) the *a priori* segmentation distribution, (2) the relationships between the atlas labels and the latent target segmentation (3) the relationships between the atlas intensities and the target intensity, and (4) the dependence / independence of the information from multiple atlases. For example, majority vote considers the segmentation estimate as the combination of equally weighted independent atlases label decisions regardless of the intensity information with a flat segmentation prior, which simplifies Eq. 25 as

$$W_{si} = \frac{\sum_j \delta(D_{ij}, s)}{\sum_l \sum_j \delta(D_{ij}, l)} \tag{2.25}$$

where $\delta(\cdot,\cdot)$ is the kronecker detla function.

Using this general framework, there are two primary categories of label fusion techniques: (1) voting label fusion attempts to find optimal weights in order to determine which atlases are optimally representative in terms of some local/semi-local/global metric [76, 77, 126, 133, 134], and (2) statistical label fusion techniques attempt to model atlas performance using a statistically driven rater performance model [125, 135-141].

5. Post-Processing

Segmented structures are generally expected to have high spatial consistency (e.g., smooth surfaces, free from speckles and pores, and etc.). Several post-processing techniques are available to provide segmentation refinement.

5.1. Markov Random Field (MRF)

The theory of MRF provides a mechanism for enforcing spatial consistency across images. Many segmentation techniques (e.g., multi-atlas label fusion [142], Bayesian methods [103, 104], Graph Cut

[69, 143]) have intermediate stages with probabilistic segmentation estimate, representing the segmentation result fuzzily. MRF can serve to regularize the voxel-wise probabilistic segmentation estimate by simultaneously considering the current segmentation estimate of the voxel itself and the neighboring voxels [107]. A neighborhood (or *clique*) structure is carefully designed to suppress the spatial inconsistency. Thus MRF provides a theoretically sound procedure to increase the consistency of the underlying segmentations.

5.2. Morphological Operation

Mathematical morphology (typically binary morphology) provides the several shift-invariant basic operations to regularize the geometrical structures of segmentation [144, 145]. A structuring element (a simple shape, e.g., a disk, square, or cross in 2-D, or corresponding geometrics in 3-D) needs to be defined so that the morphological operations take effect on how the structuring element fits or misses the underlying binary segmentation.

Specifically, erosion turns a foreground voxel into background when the structuring element fits the voxel as its center, but does not have an entire match with the local binary image. The erosion of a binary image usually "peels" one layer of the foreground object, and diminishes thin lines and speckles. On the contrary, when having a miss with the center matched up for the given structuring element, dilation turns the background into foreground on the voxels on the binary image corresponding to the missing regions of the structuring element. Dilation works for inflating foreground objects, where gaps and pores might be filled. Opening (i.e., erosion followed by dilation) and closing (i.e., dilation followed by erosion) can clear speckles and pores, respectively while keeping the most other parts of the binary image the same.

6. Performance Validation Criteria

The validation of segmentation performances is very necessary to evaluate the quality of specific segmentation approaches. Qualitatively, the segmentation results can be overlaid on the original image (in

the form of a solid region, or an outline surface) to see how the segmentation fits the underlying structure. This yields a straightforward representation of how the segmentation method performs on the tested data. More objective validation requires quantitative criteria, based on which the dis-similarities of the segmentation result to the ground truth (usually manually segmented by expert) are examined, and thus different segmentation approaches can be compared against each other statistically. Generally, volumetric accuracy and surface error are two main categories of quantitative validation criteria for medical image segmentation.

6.1. Volumetric Accuracy

Volumetric accuracy criteria compare the segmentation result against the ground truth voxel-by-voxel, collect the numbers of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), and combine them in specific manners to represent the overlapping degrees between the segmentation and truth.

Consider A as the segmentation result, B the truth volume, and $|\cdot|$ the L^1 norm operation, Dice Similarity Coefficient (DSC, [146]) is the most commonly used accuracy metric in recent studies of image segmentation.

$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \equiv \frac{2|TP|}{2|TP| + |FP| + |FN|} \tag{2.26}$$

Jaccard index is another commonly used metric for segmentation accuracy, where

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|} \equiv \frac{|TP|}{|TP| + |FP| + |FN|} \tag{2.27}$$

Sensitivity and specificity are two well-known statistical measures of performances, where

$$Sensitivity(A,B) = \frac{|A \cap B|}{|B|} \equiv \frac{|TP|}{|TP| + |FN|} \tag{2.28}$$

$$Specificity(A,B) = \frac{|\bar{A} \cap \bar{B}|}{|\bar{B}|} \equiv \frac{|TN|}{|FP| + |TN|} \tag{2.29}$$

however they are occasionally used for segmentation validation because sensitivity ignores FP and thus allows over-segmentation, and specificity focus on TN which is not a discriminative measure when the image background is large.

6.2. Surface Error

Surface error criteria characterize how far the surfaces of the segmentation and truth are from each other. These metrics are especially important when the segmentation and further clinical interests lie on the target surface rather than the target volume. Vertices are collected from the surfaces of both the segmentation and truth, based on which distances between the sets of vertices are measured in terms of their spatial coordinates. We define the vertices on the segmentation and truth surface as X and Y, respectively, and $d(\cdot,\cdot)$ as an indicator of distance measure. Then typically, mean surface distance (MSD) error and Hausdorff distance (HD) error from the segmentation to the truth can be measured as below.

$$MSD(X,Y) = \underset{y\in Y}{avg}\ \underset{x\in X}{inf}\ d\,(X,Y) \tag{2.30}$$

$$HD(X,Y) = \underset{y\in Y}{sup}\ \underset{x\in X}{inf}\ d\,(X,Y) \tag{2.31}$$

where *sup* represents the supremum, *inf* the infinum, *avg* the average. Note that MSD and HD can be measured symmetrically (calculate once from X to Y, and the other from Y to X, and then get the average) to provides more robust measures on surface errors.

PART 2

ABDOMINAL WALL

In this part, we investigate the abdominal wall from the perspectives of (1) clinical significance, (2) visual representation, and (3) automatic segmentation. The abdominal wall is not only a protective boundary for all internal abdominal organs, but also a critical structure to characterize some compound abdominal diseases, such as ventral hernia. We approach the problem by proposing a framework to provide objective criteria for clinical ventral hernia analyses (**Chapter III**). Within the framework, robust manual labeling protocol has been designed to guide the automated processes in further study including the segmentation of abdominal wall, and the derivation of hernia-related characteristics. Next, we explore the impact of the marriage of image segmentation and virtual reality (**Chapter IV**). We design a fully interactive interface within immersive virtual reality environment, and combine surface rendering of abdominal wall segmentation with volume rendering to enhance the visualization of abdomen CT for better understanding of the anatomic structures. Then, we focus on the automatic segmentation of outer surface of abdominal wall on a retrospective cohort of ventral hernia patients (**Chapter V**). Specifically, we integrate texture analysis into level set approach to improve the robustness of segmentation given the commonly observed irregular structures induced by ventral hernias. And finally, we propose an automatic framework to segment both outer and inner surfaces of the abdominal wall ranging from xiphoid process to pubic symphysis (**Chapter VI**). An augmented active shape model system integrating multi-atlas label fusion and level set techniques is presented to improve the segmentation robustness and enable accurate subcutaneous/visceral fat measurement.

CHAPTER III

QUANTITATIVE CT IMAGING OF VENTRAL HERNIAS: PRELIMINARY VALIDATION OF AN ANATOMICAL LABELING PROTOCOL

1. Introduction

Ventral abdominal hernia (VH) repair is one of the most commonly performed general surgery procedures worldwide. In the United States, nearly 350,000 repairs are performed annually at an estimated total direct cost of $3.2 billion [147]. Despite the frequency of VH repair, failure rates are high, with recurrence rates estimated at between 24 and 43 percent [5]. One possible reason for these suboptimal outcomes is a lack of evidence on the most appropriate surgical approach for different patients. At present, decisions on laparoscopic versus open repair, mesh type, mesh position, and method of mesh fixation are typically driven more by the surgeon's personal preference than by objective data [148, 149].

Multiple factors impact the success of VH repair. These factors include preoperative conditions (e.g., obesity, nicotine use, previous infections), hernia characteristics, operative technique and perioperative care (e.g., perioperative antibiotics, operative time). To date, there is no standardized method for VH classification that consistently and efficiently describes hernia characteristics. The most well-known VH classification system is the European Hernia Society Classification for Ventral Hernia (EHSCVH) [1]. This manual system can be cumbersome to use, has been unevenly accepted by surgeons, and is inconsistently applied, especially for complex hernias. The EHSCVH is semi-quantitative in that VH's are classified based on categorical locations with limited direct assessment of hernia size. The Ventral Hernia Working Group (VHWG) proposed a hernia grading system to access patients' risk for surgical-site occurrences based on more comprehensive clinical factors of patients and wounds [150];

however this classification system and its variant [151] are also not commonly used given that the involved factors are complicated to access. We hypothesize that a quantitative imaging approach will provide a more objective, efficient, and reproducible means of describing VH, and that this approach may inform future evidence-based research to improve VH repair outcomes.

The purpose of this paper is to present a standardized method for quantitative anatomical labeling of VH using standard of care computed tomography (CT). We propose a detailed anatomical labeling protocol to capture the clinically relevant geometric properties of both VH and the abdominal wall. Then, using a test dataset of human subjects with VH, we demonstrate both intra- and inter-rater reproducibility of our labeling protocol for generating key quantitative descriptive parameters, including VH volumes and the relationship of VH to relevant anatomical landmarks. We perform preliminary statistical tests on the ability of the derived VH properties to predict a relevant clinical endpoint (requirement for mesh bridge closure during VHR), with comparisons to predictions from EHSCVH metrics. In discussion, we present the main contributions of our approach and its potential clinical impact, compare its practical efficacy with other related efforts on VH characterization [2, 3, 152], and envision the future work.

2. Methods

2.1. Ethics Statement

All clinical data was collected from the Vanderbilt electronic medical records systems under institutional review board approval. The full name of the institutional review board is Vanderbilt Human Research Protection Program. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5). Written informed consent was obtained from all patients for being included in the study. An addendum was also obtained for the specific use for this study.

Table III.1. Clinically relevant quantitative parameters for describing VH.

Category	Example quantitative parameters	Description	Significance
Location	Point distance from xiphoid process, umbilicus, linea alba, linea semilunaris, ASIS, and pubic symphysis	Relative location of VH with respect to bony landmarks and fascial boundaries	Reference for hernia classification [1]
Size/shape	Defect area	Area of abdominal wall fascial defect	Critical to selection of surgical techniques for hernia repair [2]
	Maximum dimensions	VH range on three orientations	Reference for hernia classification [1]
	$V_{hernia}/V_{abdomen}$	Ratio of volume size between hernia sac and abdominal cavity	A normalized indicator of hernia severity [3]
Mechanical	Compliance[a]	Ability of muscular tissues to yield elastically on a force	An indicator that correlates with ease of repair and recurrence rate [4]

[a] Not currently accessible via CT imaging.

2.2. Quantitative anatomical description of VH

Since our objective was to create a comprehensive anatomical description of VH using quantitative parameters derived from CT, we began by assembling a set of quantitative parameters that would be relevant for informing clinical decision-making on VH repair (**Table III.1**). These quantitative parameters are divided into those describing location of the VH relative to key anatomical landmarks (including the xiphoid process, umbilicus, linea alba, linea semilunaris, anterior superior and inferior iliac spines, and pubic symphysis), those describing size and shape characteristics of the VH itself (including hernia volume, ratio of hernia volume to abdominal cavity volume [3], and defect area), and those describing mechanical properties (chiefly compliance) of the abdominal wall.

2.3. Labeling protocol

We then designed a standardized anatomical labeling protocol to enable algorithmic calculation of the above parameters (**Figures III.1, III.2**). The protocol was created for manual implementation by a research associate with experience in anatomical labeling but without specific experience in either abdominal radiology or general surgery; the protocol is also flexible enough to be used as a foundation for future semi-automated or fully automated approaches. The detailed protocol is provided in the supporting information. Briefly, it involves the following steps: (1) Select the axial and sagittal image slices on which to label the abdominal wall. (2) Label the anterior and posterior abdominal wall, and label the linea

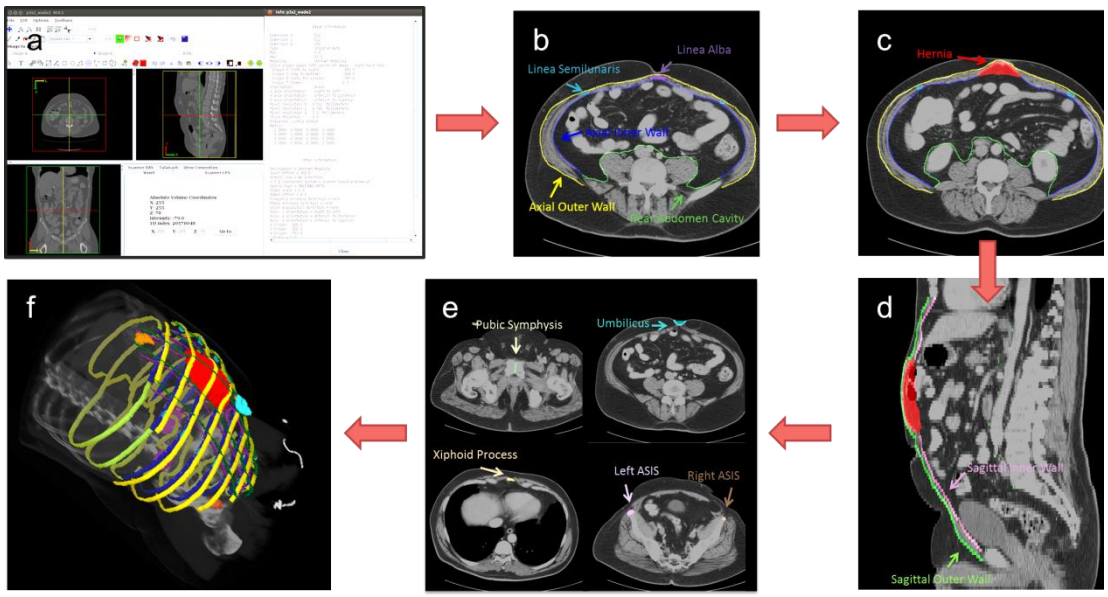

Figure III.1. Overview of the anatomical labeling protocol. (a) Axial and sagittal slices to label are determined in terms of the size and resolution of the volume. (b) On the selected axial slices, the anterior (outer and inner borders) and posterior abdominal wall is traced. At the same time, linea alba and linea semilunaris are labeled on the appropriate axial slices. (c) The VH is labeled entirely on every axial slice where the hernia exists. (d) On the selected sagittal slices, the outer and inner borders of the anterior abdominal wall are traced. Note the previous VH and abdominal wall labels can be helpful references. (e) The umbilicus and skeletal landmarks are labeled. (f) The complete set of labels is reviewed.

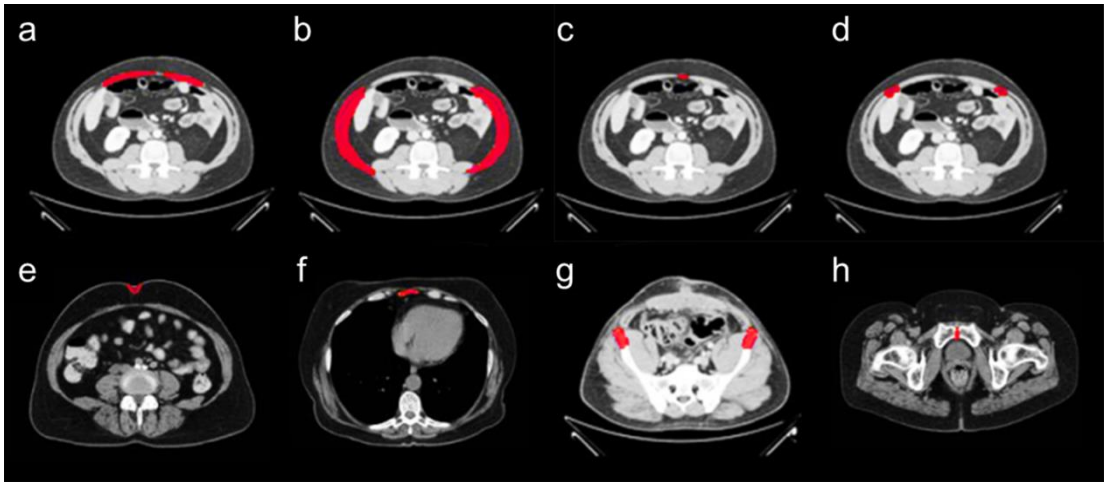

Figure III.2. Anatomical structures included in the CT labeling protocol. (a) rectus muscles; (b) oblique abdominal muscles; (c) linea alba; (d) linea semilunaris; (e) umbilicus; (f) xiphoid process; (g) anterior superior iliac spines; and (h) pubic symphysis.

alba and linea semilunaris on the appropriate axial slices. (3) Label the herniated region entirely. (4) Label the anterior abdominal wall on selected sagittal slices. (5) Label the skeletal landmarks and the umbilicus. (6) Review the overview of all labels. For manual labeling by a trained research associate, the entire labeling process takes approximately 1 hour for a complete abdominal CT volume.

2.4. Data

Retrospective, clinically acquired CT data on 61 patients with suspected VHs were collected anonymously under institutional review board approval. Abdominal scans (covering from xiphoid process superiorly to pubic symphysis inferiorly) were available for the 61 patients. 18 patients were randomly selected for protocol development; an additional random 43 patients were included for a preliminary quantitative evaluation based on the protocol. Large variations were seen among the volumes in voxels (512x512x90 ~ 512x512x200) and resolution (0.6x0.6x5 mm ~ 1.0x1.0x3 mm). Average field of view in millimeters was approximately 400x400x500 mm. Various sizes of VH were observed among the involved patients (**Figure III.3**).

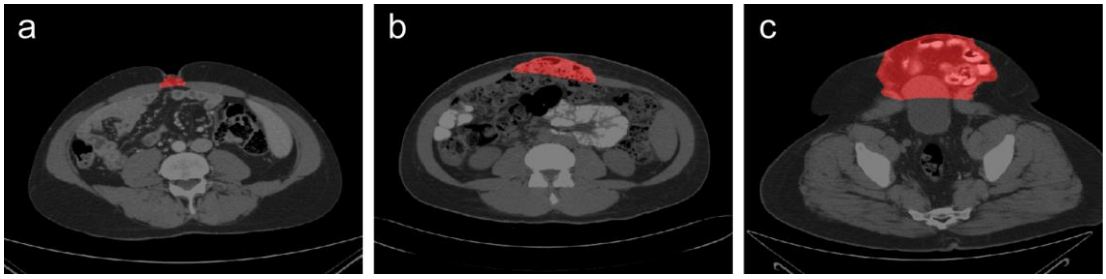

Figure III.3. Examples of various ventral hernia sizes. (a), (b), (c) demonstrate a small, medium, and large hernia, respectively in axial slices. The herniated regions are highlighted in red.

2.5. Manual labeling

A research associate was trained on the protocol using the Medical Image Processing And Visualization (MIPAV) [153] software (National Institutes of Health, Bethesda, MD) and a high resolution tablet input (Wacom, Tokyo, Japan) on a 64-bit Linux workstation. The research associate labeled all 61 datasets, with 18 of them labeled twice in randomized order with a minimum of 3 weeks between repeated volumes to ensure washout. All labels were created independently so that the research associate could not see his own prior labels. For efficiency, normal wall anatomy was evaluated on slices spaced every 5 cm. Labeling time ranged between 60 and 90 minutes per dataset. A general surgeon was also trained on the protocol and labeled a randomly selected subset of 10 of the 18 datasets. Independently, a general surgeon applied the EHSCVH protocol to the 26 patients who underwent surgical repair.

2.6. Protocol validation

Intra- and inter-rater reliabilities were estimated from the differences between the paired results for labeling the abdominal wall, key anatomical landmarks, and the VH. For the abdominal wall, reliability was calculated using the mean surface distance (MSD) and Hausdorff distance (HD) between the two sets of labels. For key anatomical landmarks, reliability was calculated with the Euclidian distance (ED) of centroids using the centermost subcutaneous point for the umbilicus; the centroids for the xiphoid process, linea alba, and linea semilunaris; the most superior point for the pubic symphysis, and the most

Table III.2. Quantitative evaluations on 20 derived metrics.

Index[a]	Metric[b]	mean [min, max]
A	Hernia volume (cm^3)	526.26 [3.99, 3946.51]
B	Hernia L-R diameter (cm)	10.69 [2.63, 30.70]
C	Hernia A-P diameter (cm)	5.11 [0.70, 17.76]
D	Hernia C-C diameter (cm)	11.64 [2.10, 30.50]
E	Hernia anterior surface area (cm^2)	313.75 [15.28, 1269.79]
F	Hernia posterior surface area (cm^2)	265.27 [10.14, 1103.02]
G	Average A-P hernia thickness (cm^2)	1.16 [0.19, 3.58]
H	Normalized horizontal hernia location	0.60 [0.25, 1.34]
I	Normalized vertical hernia location	0.16 [-0.08, 0.65]
J	Distance from hernia to left ASIS (cm)	15.30 [2.24, 24.87]
K	Distance from hernia to right ASIS (cm)	16.44 [5.16, 25.12]
L	Distance from hernia to XP (cm)	23.91 [8.89, 42.92]
M	Body volume over abdomen (cm^3)	30331.23 [19057.96, 57735.54]
N	Abdominal cavity volume (cm^3)	8809.37 [4384.98, 18611.35]
O	Ratio of hernia to abdominal cavity volume	0.06 [0, 0.41]
P	Mean abdominal wall thickness (cm)	1.38 [0.76, 2.58]
Q	Std. of abdominal wall thickness (cm)	0.80 [0.33, 2.30]
R	Visceral fat volume (cm^3)	2073.50 [5.94, 10322.78]
S	Subcutaneous fat volume (cm^3)	16528.55 [296.69, 33288.25]
T	Evaluated height of abdominal region (cm)	34.46 [25.00, 40.50]

[a] Each index represents its corresponding metric in a simpler form. A-G are considered as shape-related, G-L as location-related, and M-T as body-related metrics.

[b] Note that (1) the quantitative values of the shape-related metrics are only collected among the subjects with identified hernias; (2) the normalized horizontal location represents the relative position from left ASIS to right ASIS, the normalized vertical location represents the relative position from the level of left and right ASIS to xiphoid process; (3) the volumetric body-related metrics are evaluated over the vertical range with labeled abdominal walls, represented as T, i.e., the evaluated height of abdominal region.

anterior points for the anterior superior and inferior iliac spines. For the VH, reproducibility of hernia volume was assessed by Cohen's kappa statistic[154]. Intra-rater reliabilities were calculated on all 18 datasets. Inter-rater reliabilities were calculated on the 10 datasets for which labels were available from both the research associate and the general surgeon.

2.7. Metrics derivation

Based on the manual labels on 61 datasets, 20 metrics were automatically derived to describe the shape, location, and surrounding environment of hernias (**Table III.2**). Generally, the hernia shapes were

directly derived from the labeled hernia volume, the hernia locations were measured from the centroids of hernia volumes to the landmarks, and the surrounding body-related metrics were calculated based on interpolated abdominal wall surfaces and segmented body masks. Thin-plate spline interpolation [155] was applied to the label meshes of the abdominal walls, which yielded an outer surface for the abdominal wall as well as a closed inner surface for the entire abdominal cavity. Fuzzy C-means clustering was used to extract the patient's body from the scan table and background based on intensity. A further intensity clustering was applied to the extracted body region to separate fat tissue from muscles after excluding the bones and air, and thus the visceral and subcutaneous fat were discriminated by the interpolated abdominal wall surfaces [32]. Note that these 20 metrics are only a subset of quantitative parameters that we are interested in this preliminary study.

2.8. Statistical tests

Of all 61 evaluated datasets, 26 patients underwent VH repair with intent for primary fascial closure and mesh sublay. In general, primary fascial closure is desirable during VHR whereby the hernia defect is re-approximated. When this cannot be achieved, a mesh bridge is required, leaving the original hernia defect in situ. In other words, a patient who fails to have the primary fascial closure after VHR meets the bridge requirement. These 26 patients can be classified into two groups, where 9 patients required a bridge for closure and the other 17 did not. A series of statistical tests were used to explore the clinical correlation between the derived metrics with the technical outcomes of fascial closure.

Firstly, unpaired one-tail t-test was used to examine the significant differences between the two groups for each of the 20 derived metrics.

Then, two complementary analyses based on elastic net regularized logistic regression were operated to evaluate the compound outcome prediction using multiple metrics. The goal is to construct a regression model, based on which provide an intercept β_0 and a set of regression coefficients $\beta \in \mathbb{R}^{p\times 1}$ associated with the p metrics to minimize the deviance of model fit to the responses given N observations.

$$\beta_0, \beta = \underset{\beta_0, \beta}{\alpha\text{rgmin}} \left(\frac{1}{N} Deviance(\beta_0, \beta) + \lambda P_\alpha(\beta) \right) \tag{3.1}$$

where the $Deviance(\cdot,\cdot)$ was computed under binomial distribution for logistic regression [156] to deal with binary categorical cases while estimating the odds in a continuous form. λ represents a non-negative regularization parameter for the penalty term $P_\alpha(\beta)$.

$$P_\alpha(\beta) = \frac{1-\alpha}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \tag{3.2}$$

where both the L1 and L2 norms of the regression coefficients were used for elastic net regularization [157] to constrain the regression, where some highly related metrics can be ignored (the regression coefficient approaches zero). The regularization effect is controlled by a parameter, i.e., alpha (or α), ranging from 0 to 1, which effectively determines the proportions of ignored regression coefficients - a larger α value leads to more metrics to be ignored. Note when performing the tests, the largest value was determined among a sequence of valid candidates for λ such that the deviance is within one standard error of the minimum, leaving α as the only variable.

Here, the 20 derived metrics were all included to construct the regression model that represents how multiple metrics are combined with regression coefficients to predict the technical outcomes. Specifically,

(1) A predictive analysis was conducted to assess the accuracy of the predictive model using a leave-one-out cross-validation (LOOCV) scheme, i.e., building a regression model on 25 observations, and then testing it on the one left over. The continuous estimated outcomes were rounded into binary predictions for validation. By iterating this cross-validated analysis over α from 0 to 1 with a step size of 0.01, the numbers of remaining metrics, and those of false predictions were recorded.

(2) A follow-up exploratory test was conducted by building the regression model on all 26 observations (without cross-validation) with α of 0.9 to identify the most predictive metrics for mesh bridge requirement after VH repair.

An additional predictive analysis using elastic net regularized logistic regression with LOOCV was conducted with the variables proposed in EHSCHV for reference. According to EHSCHV, four categorical variables, i.e., medial hernia location, lateral hernia location, recurrence of hernia, and hernia width categories, and two quantitative variables, hernia width (L-R diameter) and length (C-C diameter) acquired by clinicians were considered in place of the 20 labeling-derived metrics for regression.

3. Results

3.1. Labeling reproducibility

Of the 18 patients evaluated for labeling reliability, mean age was 50 years with 50 percent women. The mean body mass index (BMI) was 33.1 kg/m^2. The prevalence of hernia in this population was 78%, and all of the ventral hernias were related to a previous operation (i.e. incisional hernias). The mean transverse dimension of the hernia defects was 11.4 cm.

3.1.1. Abdominal wall

Our protocol yielded high intra-rater and inter-rater reproducibility for labeling the anterior abdominal wall, with MSDs of around 2 mm and HDs of around 30 mm (**Table III.3**). There was moderate reproducibility for the posterior abdominal wall, with an intra-rater MSD of 2.5 mm and an inter-rater MSD of 7.7 mm. The different choices on the starting and ending point of structure labeling cause the relatively large values of HD (up to 9 cm), while the MSD values are not greatly affected.

Table III.3. Abdominal wall reliability measured by mean surfaces distances (MSD) and Hausdorff distance (HD) in mm.

MSD	Axial Outer	Axial Inner	Sag. Outer	Sag. Inner	Rear Abd.
Intra-rater	1.24 ±1.34	1.18 ±1.62	0.96 ±0.49	1.23 ±1.21	2.47 ±1.37
Inter-rater	1.16 ±0.46	1.34 ±0.73	1.93 ±1.91	2.14 ±1.47	7.73 ±2.58
HD	**Axial Outer**	**Axial Inner**	**Sag. Outer**	**Sag. Inner**	**Rear Abd.**
Intra-rater	19.93 ±17.05	16.28 ±16.17	19.84 ±13.66	20.55 ±22.49	36.17 ±24.30
Inter-rater	16.81 ±10.88	22.85 ±19.82	33.60 ±18.25	36.36 ±22.74	90.36 ±9.74

Table III.4. Fascial boundaries and bony structures reliability measured by Euclidean distance (ED) of centroids in mm.

ED	Xiphoid Process	Left ASIS	Right ASIS	Umbilicus
Intra-rater	2.61 ±3.93	3.00 ±1.99	2.30 ±1.81	1.88 ±1.89
Inter-rater	2.43 ±1.84	3.51 ±3.15	1.83 ±1.39	4.69 ±5.91
ED	**Linea Alba**	**Linea Semilunaris**	**Pubic Symphysis**	
Intra-rater	4.59 ±5.91	4.38 ±3.04	1.16 ±1.15	
Inter-rater	3.59 ±2.65	4.55 ±1.20	2.51 ±1.74	

3.1.2. Key anatomical landmarks

Reproducibility of labels for key anatomical landmarks was moderate to high, with both intra-rater and inter-rater Euclidian distances below 5 mm (**Table III.4**).

3.1.3. Hernia volumes

Reproducibility of hernia volumes was high, with an intra-rater Cohen's kappa of 0.8 and an inter-rater Cohen's kappa of 0.9 (**Table III.5**).

3.2. Quantitative evaluation

Of the 61 patients evaluated for metrics derivation, the prevalence of hernia was 72%. The mean age was 52 years with 66 percent women. The mean BMI was 33.0 kg/m^2. The mean and range of 20 derived metrics were calculated (**Table III.2**).

Table III.5. Hernia volume reliability.

Comparison	Cohen's kappa
Intra-rater	0.83 ±0.05
Inter-rater	0.92 ±0.02

Table III.6. Statistical comparison of 20 metrics between two groups of patients with distinct outcomes.

Index[a]	*p*-value	mean [min, max] (required bridge closure)	mean [min, max] (no bridge required)
A	0.0001*	823.56 [57.8, 2300.04]	107.65 [3.99, 367.81]
B	0.0000*	15.74 [8.51, 24.15]	7.20 [2.80, 12.80]
C	0.0029*	6.27 [3.05, 10.46]	3.60 [0.70, 7.24]
D	0.0004*	16.99 [4.50, 26.70]	8.14 [2.40, 22.20]
E	0.0000*	512.22 [78.93, 1036.63]	123.24 [16.30, 283.31]
F	0.0000*	448.13 [64.13, 1022.69]	105.59 [14.98, 271.81]
G	0.0001*	1.40 [0.73, 2.22]	0.72 [0.19, 1.61]
H	0.2955	0.63 [0.33, 1.01]	0.68 [0.25, 1.34]
I	0.3980	0.10 [0, 0.19]	0.09 [-0.02, 0.24]
J	0.3704	16.46 [2.24, 22.84]	15.83 [8.03, 24.87]
K	0.1834	18.54 [11.37, 25.12]	16.91 [5.16, 23.72]
L	0.4817	19.66 [13.00, 30.48]	19.55 [8.89, 30.96]
M	0.1416	32384.43 [22725.75, 42339.98]	29078.13 [19727.07, 46436.02]
N	0.4412	9233.44 [6007.36, 15065.04]	9074.89 [4683.12, 15141.70]
O	0.0005*	0.09 [0.01, 0.31]	0.01 [0, 0.04]
P	0.0539	1.55 [0.83, 2.44]	1.31 [0.76, 1.62]
Q	0.0083*	0.98 [0.39, 1.89]	0.65 [0.33, 1.07]
R	0.0655	32.48 [7.12, 65.90]	23.11 [5.94, 48.88]
S	0.1794	22960.45 [12114.26, 33288.25]	20346.47 [8218.99, 31283.60]
T	0.3847	35.68 [33.00, 38.00]	35.39 [30.30, 39.90]

[a] Each index refers to a metric in Table III.2.

* indicates significant difference between the two groups.

3.3. Clinical correlation

Of the 26 patients who underwent VH repair with intent for fascial closure and were evaluated for statistical tests, the mean age was 51 years with 81 percent female with 32.1 kg/m^2 mean BMI.

3.3.1. Unpaired one-tail t-test

Significant differences ($p<0.05$) were observed between two groups of patients separated by bridge closure requirement after VH repair over nine individual metrics (**Table III.6**).

3.3.2. Predictive regression analysis

With cross-validation, the regression model based on labeling-derived metrics and the one based on EHSCHV variables were tested. In neither case, perfect prediction was achieved. For the best case along different selection of the alpha value, the regression model using the labeling-derived metrics yields

four false predictions out of 26 subjects (84.6% accuracy), where two ("hernia L-R diameter" and "hernia anterior surface area", with $\alpha = 0.95$ or 1.00) to five ("hernia L-R diameter", "hernia C-C diameter", "hernia anterior surface area", "hernia posterior surface area", and "average A-P hernia thickness", with $\alpha = 0.80$ or 0.89) metrics were used. On the other hand, the regression model using the EHS variables made at least six false predictions (76.9% accuracy), where five (except for the hernia width category) to all six variables were included (**Figure III.4a, 4b**). With a closer look, all false predictions mentioned above were the subjects who required bridge repair but predicted as not required. These false predictions were confusing in terms of their similar hernia sizes to those having primary fascial closure after VHR.

3.3.3. Exploratory regression analysis

Only two metrics ("hernia L-R diameter" and "hernia anterior surface area") remained as the key factors for predicting mesh bridge requirement when building a logistic regression model with an alpha value of 0.9. This regression model, although used all observations, yielded the same four false estimates

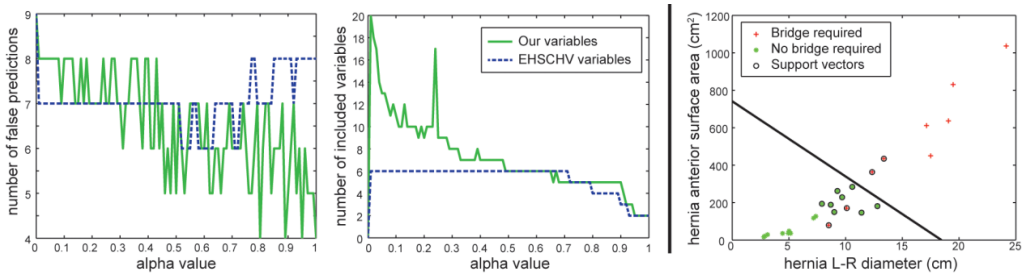

Figure III.4. Results of preliminary statistical analyses. (a) and (b) shows the number of false predictions and number of included variables over different alpha values using cross-validated elastic net regularized logistic regression, respectively. Generally, a larger alpha value yields stronger regularization, and thus involves less variables for the regression model. Note that the blue dashed curves represent the regression results using EHSCHV variables, while the green solid curves use the variables derived from labeling. (c) presents a hyper-plane using support vector machine to separate the two groups of patients with distinct technical outcomes by the two remaining labeling-derived variables of an exploratory regression model built upon all observations.

as the best cases using labeling-derived metrics in the predictive analysis. These two metrics can be used to identify a separating hyper-plane (a discriminative function of the two metrics) between two groups using support vector machine (SVM) [158], which represents the quantitative threshold for mesh bridge requirement. Two cases requiring bridge repair were misclassified (**Figure III.4c**). We note that the uneven numbers between the two groups (9 vs. 17) can affect the regression model regarding its optimal threshold (we use 0.5 for logistic regression) for making predictions. Applying a SVN classifier following the regularized regression model may raise the predictive power.

4. Discussion

4.1. Main contributions

This preliminary study is intended to lay the foundation for a quantitative imaging approach to determine optimal management strategies for different subtypes of VH, and improve the surgical treatments.

First, we created a set of clinically relevant, quantitative anatomical descriptors for VH, and we designed a standardized labeling protocol to enable extraction of these parameters from routine clinical CT datasets as the foundation for future automated modeling of the relationships between VH anatomical characteristics and treatment outcomes.

Second, our protocol validation study showed acceptable inter-rater and intra-rater reproducibility for labeling the abdominal wall, key anatomical landmarks, and the VH itself. In terms of anatomical labeling, we found that abdominal wall surfaces could be appreciated on either axial or sagittal views without extensive three-dimensional visualization, while the hernia volume required tri-planar manipulation. We also found that abdominal wall surfaces are sufficiently smooth such that, for efficiency, normal wall anatomy can be labeled every 5 cm on sparse, evenly spaced slices; the entire surfaces can then be approximated by interpolation.

Next, we derived 20 quantitative parameters to describe the shape, location, and surrounding environment of VH from the anatomical labeling automatically. The collection of these metrics provide more comprehensive characteristic of VH than the available clinical measurements using EHSCHV system.

Lastly, we showed the clinical correlation between the derived quantitative parameters and the technical outcomes of primary fascial closure after VH repair with preliminary statistical tests. 9 individual metrics were shown to be significantly different between patients required bridge closure and those who did not. Through predictive analyses, we presented a regression model using multiple metrics that were capable of identifying all patients who did not require bridge closure (17 out of 17), and over half of the patients who required (5 out of 9). We also found our labeling-derived metrics more predictive than EHSCHV variables for this technical outcome. We note this is the first work to correlate VH quantification into clinically meaningful disease processes.

4.2. Potential clinical impact

Clinically, the requirement of bridging remains uncertain pre-operatively. Surgeons may give a rough prediction by eyeballing the hernia on pre-operative imaging, i.e., a larger hernia is more likely to require a mesh bridge repair. Subjective predictions can be inaccurate. Because of this uncertainty, some techniques like myofascial release are used to avoid bridging, but may cause other problems. For example, data from our institution suggests that the extra dissection of a myofascial release significantly increases the risk of surgical site infection post-operatively [159]. Thus, objective criteria for accurately predicting the bridge requirement can be clinically relevant, and change the clinical procedure significantly. For pre-operative planning, surgeons can provide objective quantities rather than subjective size description when counseling patients and planning operations. The patient and surgeon could have an estimated probability of the need for bridge closure, and this option could be weighed while taking into account other patient factors which increase the risk of postoperative infectious complications. This could also result in lifestyle modifications prior to embarking upon a surgical repair in order to decrease the

morbidity of the procedure. In some instances, the knowledge that a bridge will likely be required in a patient who has significant risk of infectious complications might lead to surgeon to accept a bridge configuration and counsel the patient accordingly.

Our labeling protocol yields reasonable objective criteria for the predication of bridge requirement. Predictive (cross-validated) statistical analysis indicates that one needs two to four metrics to yield the best prediction of the bridge requirement. This suggests that multiple metrics should be considered together for prediction. In explanatory analysis (not cross-validated), only two variables ("hernia L-R diameter" and " hernia anterior surface area ") are needed to identify a separating hyper-plane (a discriminative function of the two variables) between the groups, which represents the quantitative threshold for the requirement of bridge repair. We note that a hernia with large L-R diameter but small anterior surface area may not necessarily require a bridge requirement, which is difficult to judge visually from 3-D CT. Therefore, these metrics are helpful as objective criteria for bridge requirement prediction. We also note that while "hernia L-R diameter" has been commonly considered as an important factor, "hernia anterior surface area" has never been focused on for VH characterization. Our labeling protocol provides the access to these parameters potentially critical to correlate specific technical outcomes. On the other hand, we find the current metrics fail to discriminate some small hernias in need of bridge repair. This is partially due to the insufficient (26) and unbalanced datasets (17 vs. 9) available in the experiment. Inclusion of more datasets could help augment the prediction of primary fascial closure. Many other factors, in addition to hernia characteristics, influence the decision to perform bridge repair. The intended goals of the operation (definitive versus staged repair), level of contamination, amount of tension on the fascia when closed, need for tissue coverage, and surgeon training can contribute to the decision-making.

4.3. Comparison to other related efforts

Our approach takes advantage of the fact that most VH patients undergo pre-surgical CT scanning to evaluate their abdomen. At present, however, no well accepted method of VH classification exists for

routine use, and therefore information from imaging is used qualitatively and subjectively to make clinical decisions based on little empirical data. The EHSCVH is the only potential classification system that has been presented as a means to classify all ventral hernias. The EHSCVH system codes for (1) categorical assignment based on the location of the hernia and cause of hernia (primary or incisional), (2) categorical assignment of hernia size (small, medium, large), and (3) linear measures of hernia size (length, width). The actual implementation of this classification system can be cumbersome, however, and although the EHSCVH was published several years ago, it has not gained widespread acceptance among surgeons. The time required of a surgeon in a busy clinical practice to determine the classification for a particular patient's hernia often prevents its use. Additionally, there is no incentive in place for the surgeon to spend the time determining the classification since the classification scheme has not been linked to specific patient outcomes or advantageous operative techniques. Finally, there are inconsistencies between surgeons when classifying complex hernias using this system. Our predictive regression analysis showed that the EHSCVH variables were not adequate to predictive the bridge closure requirement after VH repair (**Figure III.4a, 4b**) despite its simplicity.

A quantitative approach is attractive for two reasons: (1) it may be implemented by a trained associate or, in the future, by a semi-automated or fully automated computer algorithm, thus offloading the time for VH classification from the surgeon; and (2) its standardized nature provides a foundation for rigorous statistical correlation against patient outcomes, both retrospectively using large clinical databases and prospectively in clinical trials. At present, however, quantitative description of ventral hernias is rather rudimentary, with transverse size being the most commonly used metric. Transverse hernia size captures very little of the actual heterogeneity of VHs. Hernia volume is also inadequate for describing VH, as two VHs of the same volume may have very different shapes and may require different surgical techniques (**Figure III.5**).

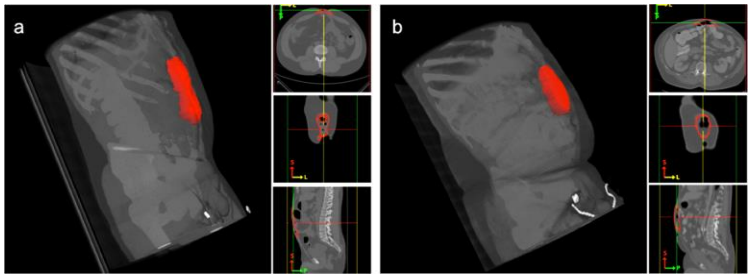

Figure III.5. Two VH cases in volume rendering and tri-planar views. Although the two examples have almost the same hernia volume size (a = 125 cm^3, b = 109 cm^3), (a) is a long, shallow rupture at the umbilicus, while (b) is a short, deep protrusion of the abdominal wall. In addition, the patients' body sizes are quite different, and the hernia in (b) is further away from the umbilicus.

For these reasons, and for its obvious clinical implications, characterization of VH has been of interest to the image processing community in recent years. Tanaka et al. derived the volumes of the hernia sac and abdominal cavity by assuming them as ellipsoid structures with the measurement of the cranio-caudal, latero-lateral, anterior-posterior radial distances [152]. Sabbagh et al. determined the intraperitoneal volumes (including both VH and abdominal cavity) after the volume boundaries were defined using a blind-side method by a surgeon and a radiologist [3]. Yao et al. marked the required range on 3-D reconstructed CT, and measured the volumes with measurement-voluminal software [2].

While notable, these efforts have not addressed the fundamental challenge of measuring the complex interaction of the hernia with its biological context. Our labeling protocol allows for estimation of different geometrical properties of VH from the labeled data (**Figures III.6, III.7, Table III.2**). We consider the hernia shapes (volumetric sizes and dimensional diameters) as the chief quantitative parameters to characterize the degree of abnormality of VH. We suggest that the location of the hernia defect with respect to bony landmarks and facial boundaries is critical to VH classification. We also append metrics of surrounding structures (e.g., abdominal wall thickness, visceral and subcutaneous fat volume) as referential body status of the VH patients. These characteristics give a robust description of the hernia itself, which can then be correlated to clinical outcomes. Elastic net regularized logistic regression can be used to reduce all available variables to several key factors to predict specific technical

outcomes, while sometimes a larger number of variables are required to yield better prediction. Certainly other non-hernia related factors are important in the overall definition of hernia complexity, including

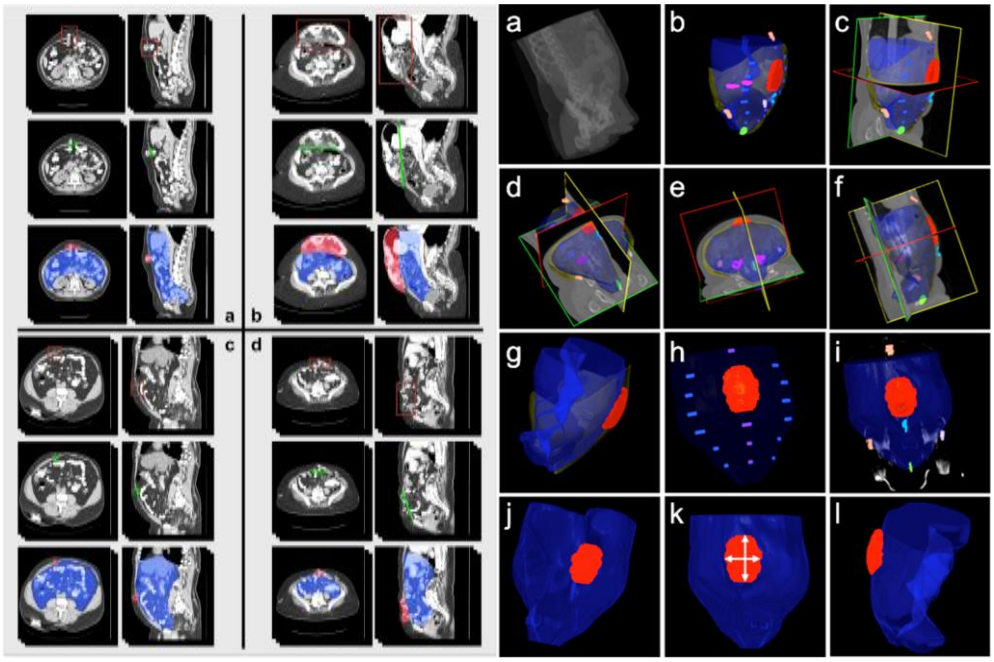

Figure III.6. (Left) Illustration of VH characteristics on CT for four patients. In each section, the first row illustrates the location of the VH; the second row illustrates the VH defect size at the anterior abdominal wall; the third row demonstrates the volume size of the hernia sac (red) and the abdominal cavity (blue).

Figure III.7. (Right) Illustration of VH characteristics in terms of processed label results. The first row, from (a) to (c), demonstrates a matchup between the original image data and the processed labels, where the abdominal walls were interpolated. The second row, from (d) to (f), demonstrates the coherence of interpolated abdominal walls with the original image in three different views. The third row, from (g) to (i), illustrates a combined model of abdominal wall and hernia volume for shape-related VH characteristics, the relative location of VH with respect to the linea alba and linea semilunaris, and the relative location of VH with respect to skeletal landmarks and the umbilicus. The fourth row, from (j) to (l), demonstrates feasibility of measuring the VH defect size, width and length of VH, and ratio of volume size between the hernia sac and the abdominal cavity.

many clinical and patient-related factors[160], but the method described in this manuscript offers a precise and reproducible description of ventral hernia based upon specific imaging characteristics.

4.4. Future work

One important property of VH not currently accessible via CT imaging is muscular compliance, i.e., the ability of muscular tissues to yield elastically to an applied force. This patient-specific property influences the ease of repair and may correlate with post-operative recurrence rates. Future studies may investigate the use of ultrasound or magnetic resonance elastography to estimate this property.

Also, we note that the CTs in our study were acquired at rest without performing a Valsalva maneuver. Valsalva maneuvers can be beneficial to show the potential visceral drift around the herniated region. In theory, the anatomical changes would not disrupt implementation of the proposed protocol. However, further study would be needed to explore the examine differences between scans with and without Valsalva maneuvers.

It is important to emphasize the clinical relevance of these findings. If a reproducible, quantitative, and automated method of classifying VH can be developed, the field of VH management could be significantly advanced. Namely, a reliable metric would be established whereby comparisons can be made to determine best practices—akin to cancer staging in the management of malignant disease. Relevant automated efforts are in progress[38].

CHAPTER IV

IMMERSIVE VIRTUAL REALITY FOR VISUALIZATION OF ABDOMINAL CT

1. Introduction

Modern medical imaging techniques produce large data sets that are difficult to visualize and understand by both medical professionals and patients. There are several difficulties that these large data sets present. First, the amount of data means that viewing axial sections becomes problematic, simply because of the large number of slices that must be navigated. The navigation modalities for these slices, such as sliders and mouse wheels, become increasingly cumbersome as the number of slices increases. Secondly, as the complexity of medical image data grows, restricting the viewing modality to a traditional two-dimensional view may be sub-optimal. Such data is inherently three-dimensional (3D), and exploring it as a 3D quantity, through displays that offer stereoscopic depth perception, may give better insights and comprehension into the overall imaging data.

Immersive virtual reality (IVR) may offer a solution to the issues described above. In this paper, we describe IVRs presented through head-mounted displays (HMDs), stereoscopic display devices that allow the wearer to perceive a three-dimensional virtual environment as though present in it. Immersive virtual reality facilitates the investigations of situations that are difficult to study in the real world, whether for reasons of cost or complexity. It can provide visualization modalities that render complex data more comprehensible and interaction modalities that make large amounts of data more accessible. In this paper, we discuss IVR that uses an HMD to view the virtual world, and articulated data gloves to interact with it.

In this paper, we consider visualization of ventral hernias. Ventral hernias occur in up to 28% of patients undergoing abdominal operations — even in optimal conditions [161, 162]. Repair of these

hernias is fraught with failure; recurrence rates ranging from 24-43% [5]. Recurrence of previously repaired VHs increases costs and morbidity to patients and can sometimes require multiple repairs. In some patients, repair of their end-stage VH may produce a worse outcome than a non-operative strategy [163]. Communication is a significant problem for clinicians and patients with ventral hernias.

In care planning, it is essential that the patient understand the urgency, degree of severity, and impact of a hernia (and potential repair) on patient quality of life. Hernias are defined by ruptures in the abdominal wall (i.e., the absence of healthy tissues) rather than a growth (e.g., cancer); therefore, understanding a hernia necessitates understanding the entire abdomen. Our proposed environment allows surgeons and patients to view body scans at scale and interact with these virtual models using a data glove.

IVRs have, of course, been used to visualize medical imaging data before. Some systems use a single imaging modality or semi-immersive environments [164-166]. Using the hands to manipulate the 3D data has also been recognized as having value [165, 167]. In particular, Indhumathi et al. [168] use both an HMD and a high-fidelity data glove to manipulate medical imaging data. The novelty of the present system is coupling of the manual interface with the IVR to achieve improved understanding of large imaging datasets.

2. Methods

2.1. Data and Processing

Our abdominal segmentation method has been evaluated in separate work [38]. Briefly, retrospective, clinically acquired CT data on three male patients and one female patient with suspected VHs (three with confirmed hernia) were acquired in anonymous form under institutional review board supervision from the clinical PACS in DICOM format. Volumes were approximately 512x512x157 voxels with a resolution of 0.89x0.89x3 mm. A low threshold (200 HU) is used to identify the whole bone skeleton. Then, a high threshold (800 HU) is used to divide the skeleton into different components based

on relative shape and position information. The skin and abdominal wall are segmented using level set techniques. The outer surfaces of each object (i.e., pelvis, femurs, spinal column, ribs, skin) were tessellated and exported for visualization. Surface colors and transparency were authored for each object using Mayavi, a 3D scientific visualization tool. These models were then exported into the virtual environment.

2.2. Materials and Apparatus

Our immersive virtual environment emulates a free-walking space (approximately 8m × 7m × 4m). The virtual environment is viewed through a full color stereo Nvis (Reston, VA) Nvisor SX60 HMD with 1280x1024 pixels per eye, a nominal field of view of 60^{o} diagonally, and a frame rate of 60Hz. An InterSense IS-900 precision motion tracker updates the user's rotational movements around all three axes, supplemented by optical tracking by four cameras of two infrared LEDs on the HMD to provide position and orientation information. The virtual environment is rendered using Vizard (Worldviz, Santa Barbara, CA). A wireless Cyberglove II data glove (Cyberglove Systems, San Jose, CA) for the right hand is used to track the fingers, enabling gesture control in our system. The data glove is instrumented with 22 high-accuracy sensors and can track the finger movements accurately. The global position and orientation of the hand is tracked with an eight camera Vicon (Los Angeles, CA) MX-F40 optical tracking system.

2.3. Mechanisms for Viewing the 3d Model the CT Slices in Immersive Virtual Environment

In the virtual environment, registered computed tomography (CT) images and the 3D model were superimposed. The CT images were made semi-transparent to make the 3D model visible and to give a volumetric effect to the ensemble. A user can interact with a given model by touching (with the index finger of the virtual hand) a CT slice to expose it. When exposed, the CT slice "pops" outside the model and becomes opaque for optimal viewing. The image can be made to disappear by brushing the image

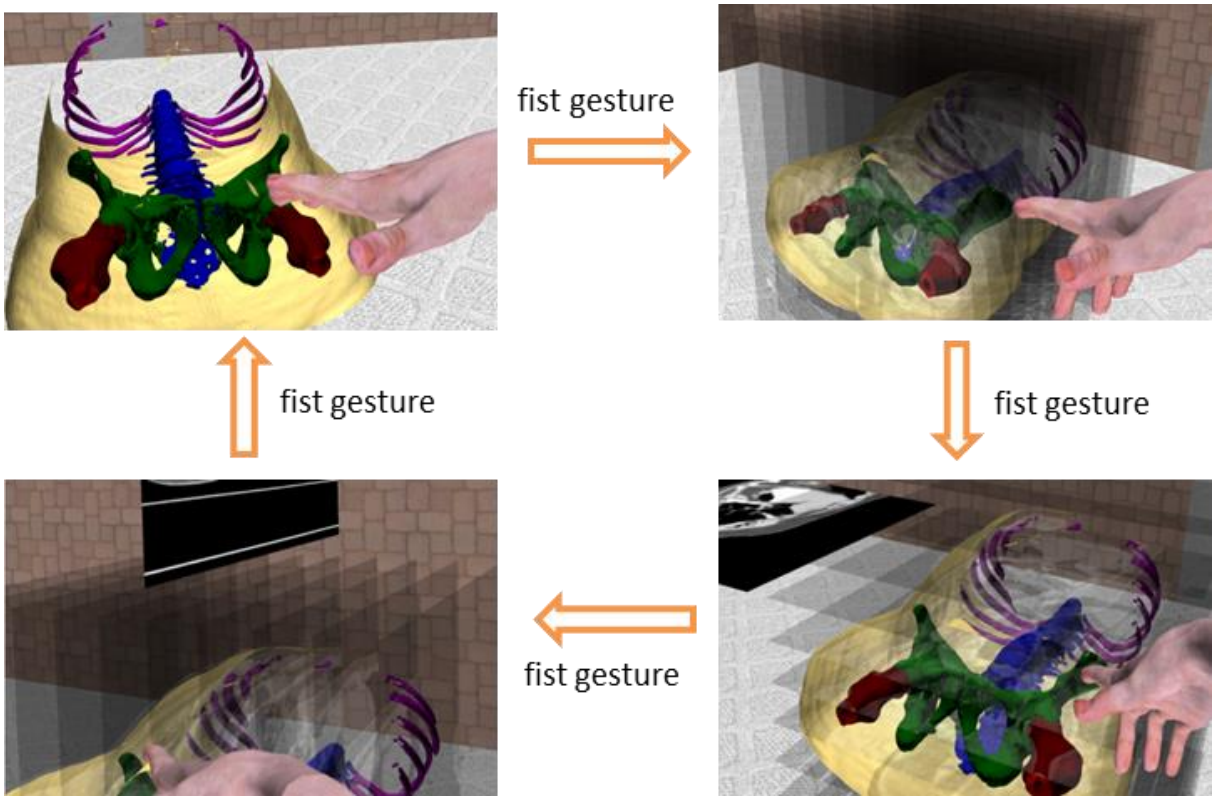

Figure IV.1. The four states of the model in IVR.

with the palm of the hand or by selecting another slice. An accordion-like affect can be achieved by running the virtual hand through the model.

Since the CT slices are along three axes, it can be difficult to select an individual slice from a complete grid. Therefore each axial direction is toggled by making a fist with one's hand. The state transition was controlled by human making a fist gesture of right hand and releasing immediately. When the fist-gesture was captured by the data glove, the system transitioned to the next state immediately. The initial state had no CT images in it; this process is shown in **Figure IV.1**. In the state with the images in the scene, a user can select the image they want to observe as mentioned in the previous paragraph.

2.4. Volume Rendering in the Immersive Virtual Environment

In scientific visualization and computer graphics, volume rendering is a set of techniques used to display a 2D projection of a 3D discretely sampled data set. Here the 3D data set is a group of 2D slice images acquired by a CT scanner. These are acquired in a regular pattern as one slice approximately every

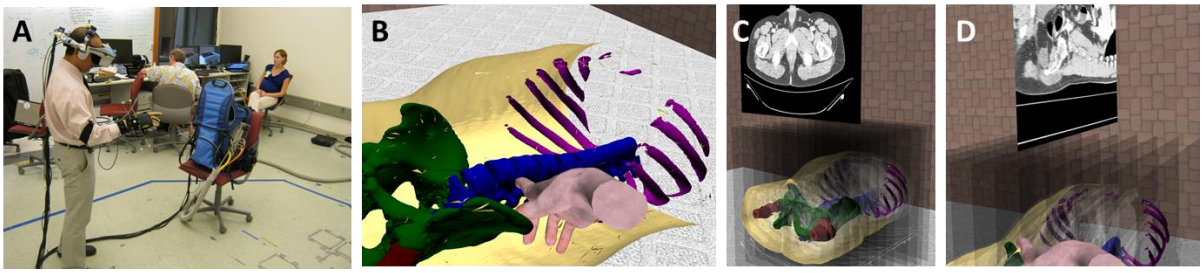

Figure IV.2. Illustrations of the proposed system in use. (a) Surgeon using the system. (b) Virtual hand interacting with 3D abdominal model. (c) Navigation of axial slices. (d) Navigation of sagittal slices.

millimeter and have a regular number of image pixels in a regular pattern. To render a 2D projection of the 3D data set, first we define a camera in space relative to the volume. The opacity and color of every voxel are also defined. Direct volume rendering is a computationally intensive task; here we achieve the volume render effect by rendering closely spaced slices with user-controlled transparency. An Xbox Controller (Microsoft, Inc., Redmond, WA) is used to adjust the mapping between intensity value (HU units) and transparency ("Alpha Histogram"), baseline brightness, baseline transparency ("alpha boost"), and to enable whole-volume intensity normalization. These controls are shown in a simulated heads-up display in the upper right field of view.

3. Results

The abdominal model contains 600893 triangles. The CT images were generated from axial (plane = left/right – front/back), coronal (plane = left/right – head/toe), and sagittal (plane = front/back – head/toe) sections through the data at every millimeter. There were 157 CT scan images in the axial direction, 458 in the sagittal direction, and 458 in the coronal direction. The number of pixels in each dimension is: 512 (left-right) x 512 (front-back) x 157 (head-toe). A pixel corresponds to a unit volume of size 0.8945 mm (left-right) x 0.8945 mm (front-back) x 3 mm (head-toe). For the slice selection interaction, ten CT slices were used in each direction.

We demonstrated the system to two surgeons using an abdominal model containing a hernia; **Figure IV.2** illustrates the system in use. Remarks from the surgeons indicate that the system was

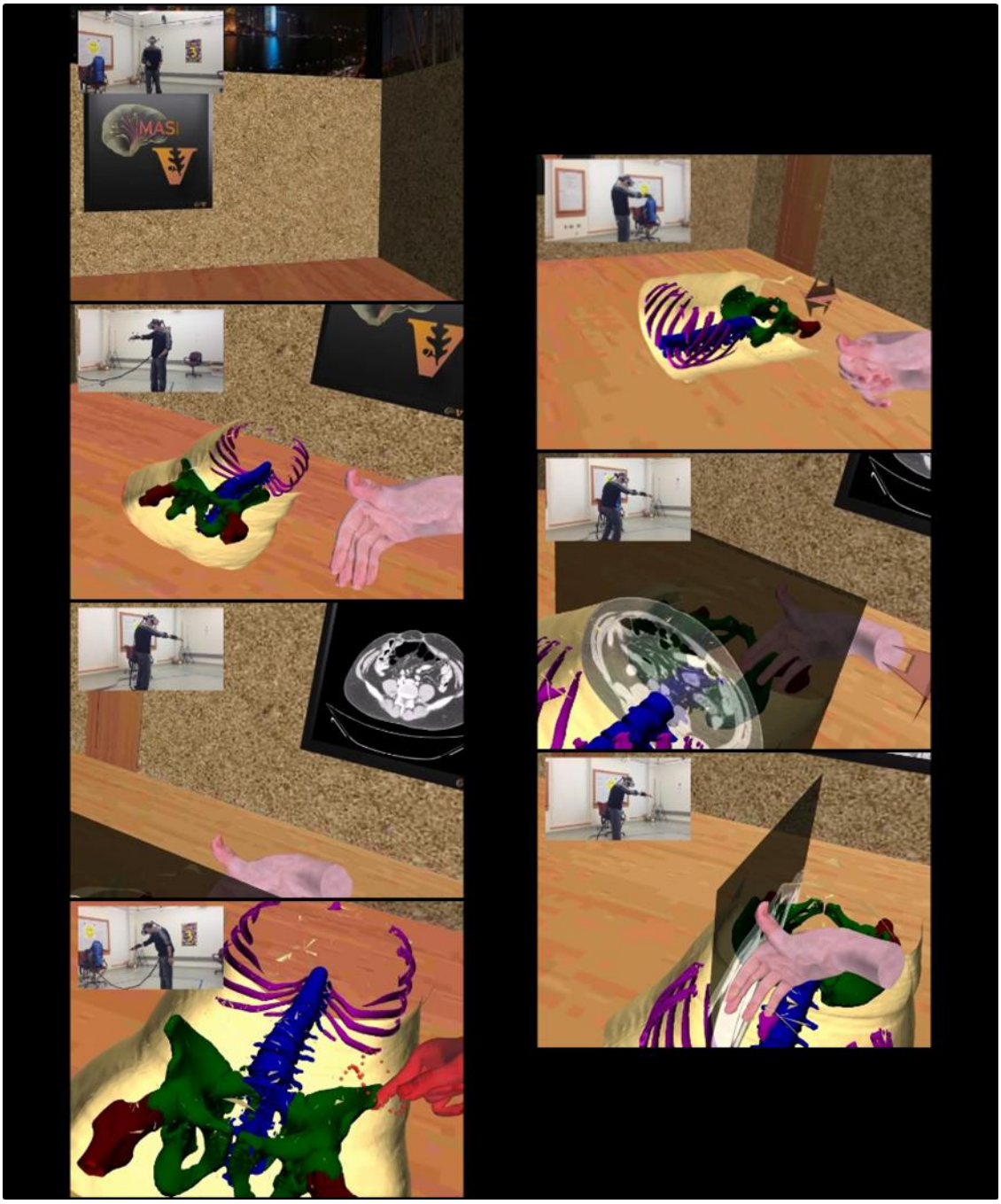

Figure IV.3. Illustration of user interaction with the abdominal model.

immersive, easy-to-use, and conveyed a sense of scale and anatomy that is difficult to achieve through a desktop display (**Figure IV.3**). They saw immediate uses for the system in both patient education and pre-operative planning.

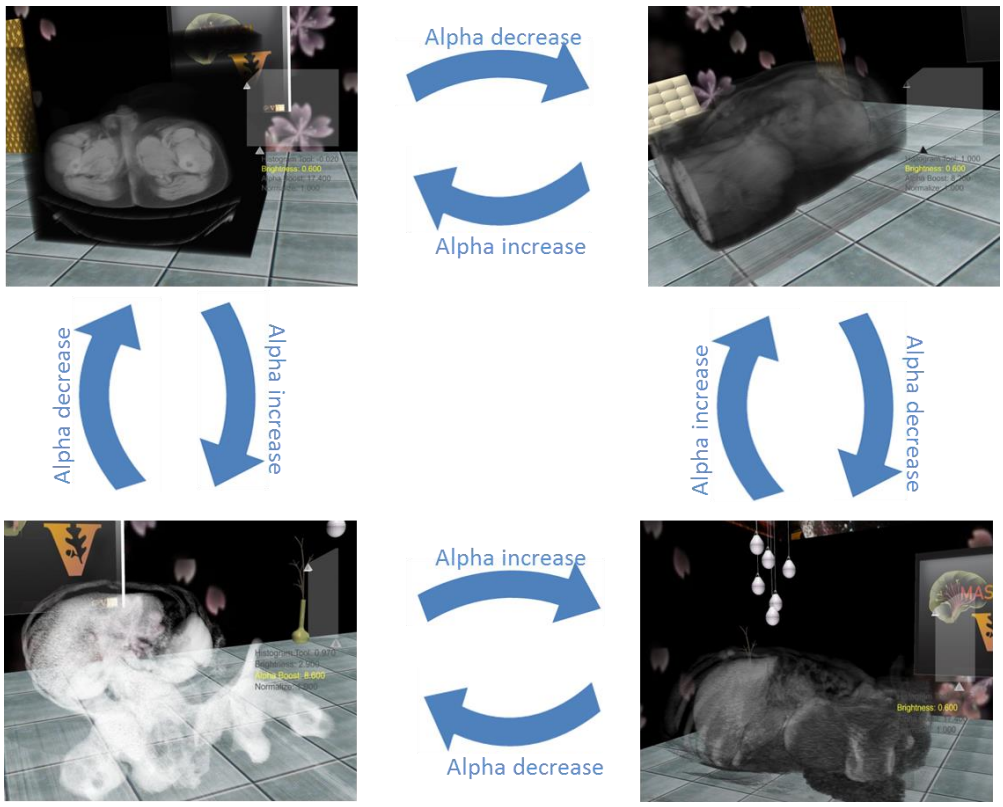

Figure IV.4. Volume rendering for abdominal wall CT scan.

The volume rendering controls enable real time interaction with the rendering properties so that users can adjust the display to reveal different parts of the anatomy from the abdominal wall to the skin, muscles, organs, and bones (as shown in **Figure IV.4**). These dynamic changes are in addition to the user's ability to walk around the model and peer over/in/through the display. Volume rendering of the CT data provides a different sense of relative anatomical relationship than the renderings of the segmented data (as in **Figures IV.1-3**).

4. Discussion

We have demonstrated a system capable of visualizing large medical imaging datasets that offers opportunities for improved understanding of such datasets. This work is a demonstration of feasibility and establishes pilot platform on which to evaluate utility of IVR in medical imaging visualization and establish the relative merits of the IVR capability. Volumetric segmentations can be viewed and interacted with alongside traditional 2D slice rendering. This combined experience preserves/creates a

perception of scale which may be non-intuitive given more traditional rendering mechanisms. The image pre-processing can be fully automated and triggered based on PACS status, so that IVR could be feasible within a clinical workflow.

Readily available flat screen virtual reality (e.g., Xbox Connect, Microsoft, Redmond, VA) and low cost IVR/HMD (e.g., <$50, Ilixco, Sacramento, CA) provide fascinating opportunities for integrating IVR technology with medical imaging systems. Important questions remain as to the relative utility of full scale (as presented, walk-around IVR) versus seated IVR versus flat screen virtual reality. Moreover, substantial work remains in optimizing intuitive interfaces for self-navigation and for guided navigation (e.g., shared experiences between patients and clinicians). Ongoing efforts are characterizing the utility of IVR for medical education and patient communication. In conclusion, integration of IVR technologies for data interaction with existing PACS infrastructures promises to be a fruitful area of exploration.

CHAPTER V

AUTOMATIC SEGMENTATION OF THE ANTERIOR ABDOMINAL WALL

1. Introduction

Ventral hernias (VH) include primary abdominal wall defects (e.g., umbilical hernias) and acquired incisional defects resulting from previous abdominal operations. The management of ventral hernias remains a challenging problem for primary care physicians, surgeons, and patients. VHs occur in up to 28% of patients undergoing abdominal operations even in optimal conditions [161, 162]. Repair of these hernias is fraught with failure; recurrence rates ranging from 24-43% are reported, even with the use of biocompatible mesh [5]. Recurrence of previously repaired VHs increases costs and morbidity to patients and can require multiple repairs. The common clinical problem of VH and wide variation in care present a unique opportunity for improvement in classification and outcomes. For each 1% reduction in recurrence of VH after repair, an estimated annual cost savings of $32 million could be realized [147].

Computed tomography (CT) is used to make qualitative clinical judgments about a particular patient's hernia for treatment and prognosis. However, the only quantitative metric currently in use is the transverse dimension of the hernia defect. We posit that the CT image obtained from these studies is underutilized and provides a potentially rich — and automated — means of better characterizing VH. Three-dimensional structural measurements based on computational tools have been the subject of extensive study in the brain in the search for biomarkers for clinical development of therapeutics [169], and have been targeted for therapeutic modifications and as surrogate endpoints in clinical trials (e.g., [14, 15]). These methods have yet to be applied to VH.

Image segmentation methods to capture the three-dimensional VH-related anatomical structures could provide a foundation on which to measure geometric properties of hernias and surrounding tissues.

For example, given the segmentation, we could numerically locate the hernia with respect to anatomical landmarks, compute the area of disruption and volume of herniated tissue, estimate the displacement and volumetric changes in abutting wall structures versus change at distant relatively normal tissues, and thus help surgeons to classify the VHs and optimize the treatment delivery. In a parallel study [34], we consider a manual labeling protocol that uses the normal appearing anterior abdominal wall and the herniated region to describe the shape-related characteristics of VH, and uses fascial boundaries and bony landmarks as the features to extract the location-related characteristics. Herein, we target reproducible automated segmentation of the outer surface of anterior abdominal wall.

There are few studies involving the segmentation of abdominal wall. Feng et al. used a 3D

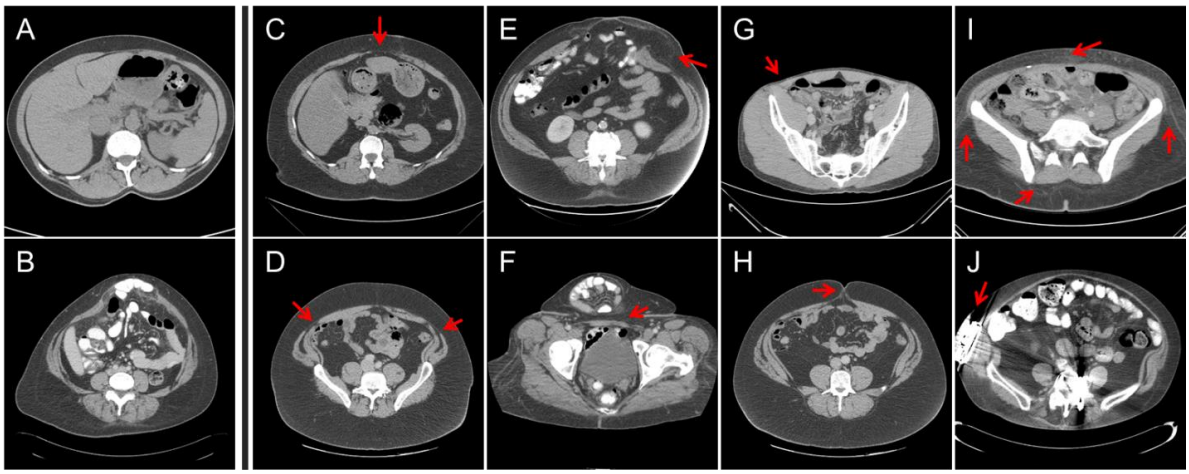

Figure V.1. Illustration of the image qualities in Ventral Hernia CT. (A) and (B) demonstrate CT scans with low artifacts for normal abdominal wall and for herniated region, respectively. (C) – (J) illustrate challenges from the segmentation of anterior abdominal wall, where red arrows indicate the challenging regions in each scenario. (C) The linea alba is thin and of lower intensity than normal. (D) The linea semilunaris is thin and of lower intensity than normal. (E) At the herniated region, the abdominal wall is stretched and can be barely seen. (F) The hernia volume is folded, which introduces a large curvature. (G) The patient is slim, which makes it hard to differentiate the muscles from the skin. (H) The umbilicus can interfere with the smooth contour of the abdominal wall. (I) Speckles in the fat are of similar intensity with muscles. (J) Metal implants result in streaking artifacts in CT scans.

flipping-free deformable model to register the inner boundary of the wall for ease of segmentation and visualization of abdominal organs [170]. Zhu et al. provided an interactive approach to remove the entire the abdominal wall to reduce the sliding motion effect on the non-rigid registration of abdominal images [33]. Yao et al. segmented the outer surface of the abdominal wall to separate subcutaneous and visceral adipose tissue by fuzzy c-means clustering and active contour models [32]. Among these studies, the abdominal wall was extracted to provide a better access to other abdominal structures, but not considered as the main entity for precise quantitative analysis. However, for our purpose of large-scale VH classification and characterization, a fully automated approach is required to provide the segmentation of a smooth surface of the anterior abdominal wall accurately and reproducibly. Level set image processing methods can be ideally adapted to finding contours between objects of different intensity in the presence of noise, artifacts, and disruption through the use of tuned regularization criteria [117]. Abnormal hernia geometries and various image artifacts can make the voxel-wise intensity information misleading, and thus cause the segmentation of the abdominal wall to be challenging (**Figure V.1**). On the other hand, the intra-tissue variability appears to form repeated patterns within local regions. These local textures are visually distinguishable between different tissues, e.g., adipose tissue looks grainy, whereas muscles and bones are dense. This poses an interesting question: can texture analysis improve the level set segmentation within the abdominal region, especially for the segmentation of anterior abdominal wall?

Texture analysis has been a long studied technique in computer vision [53], and has been applied to medical imaging [54]. Statistical methods (e.g., co-occurrence matrices [55, 56]) are used if the texture is considered as the spatial distribution of intensities, where second-order statistics of the image is estimated. Geometrical methods (e.g., Voronoi tessellation [171]) identify building elements for texture, and then assess the statistical properties of these elements, or extract the placement rule that characterizes the texture. Model-based methods (e.g., Markov random fields [172], fractals [173]) establish image models that capture essential qualities of texture for ease of texture description and synthesis. Filter-based methods extract features in frequency domain by passing images through multi-channel filter operators. Gabor filters, for example, are capable of extracting frequencies and orientations at multiple scales from

images as texture features for classification and segmentation [58-62]. Several studies have integrated texture analysis with level set techniques [63-65], among which Paragios et al. proposed a sophisticated geodesic active regions framework that combines both boundary- and region-based modules extracted from texture analysis into one level set objective function for supervised texture segmentation [174] — these studies mainly targeted partitioning regular textures (wood, fabric, zebra, leopard, etc.). Here, we assess the feasibility of the perceived local textures (as an alternative of/in addition to the intensity values), to drive more effective level set segmentation for anterior abdominal wall. We propose a direct approach to use texture analysis for level set segmentation. In particular, we learn texture features by Gabor filters, cluster the features into voxel-wise probability membership, and guide level set evolution by the local differences of the membership in place of original intensity values. Segmentation results of the proposed approach are compared against the methods only using intensity information.

We note that our interest in the abdominal wall is motivated by our desire to study VHs. Therefore, our population of interest includes subjects with poor imaging contrast, abnormal anatomy, metallic implants, etc. The existence of VHs greatly challenges the segmentation of abdominal wall (**Figure V.1**). We are evaluating the performance of our image processing methods in the context of these considerations to preserve robustness. These issues motivate the potential clinical impact of this work.

2. Theory

Herein, we describe the details on the algorithm and implementation of the proposed texture-segmentation of abdominal wall on ventral CT scans with suspected VH. We provide brief description of the segmentation of bony skeleton and skin surface as other essential structures for VH characterization (**Figure V.2**).

2.1. Algorithm

The anterior abdominal wall is formed by muscles (e.g., rectus, obliques) and fascial connections (e.g., linea alba, linea semilunaris) encompassing the abdominal cavity. Given the distinguishable

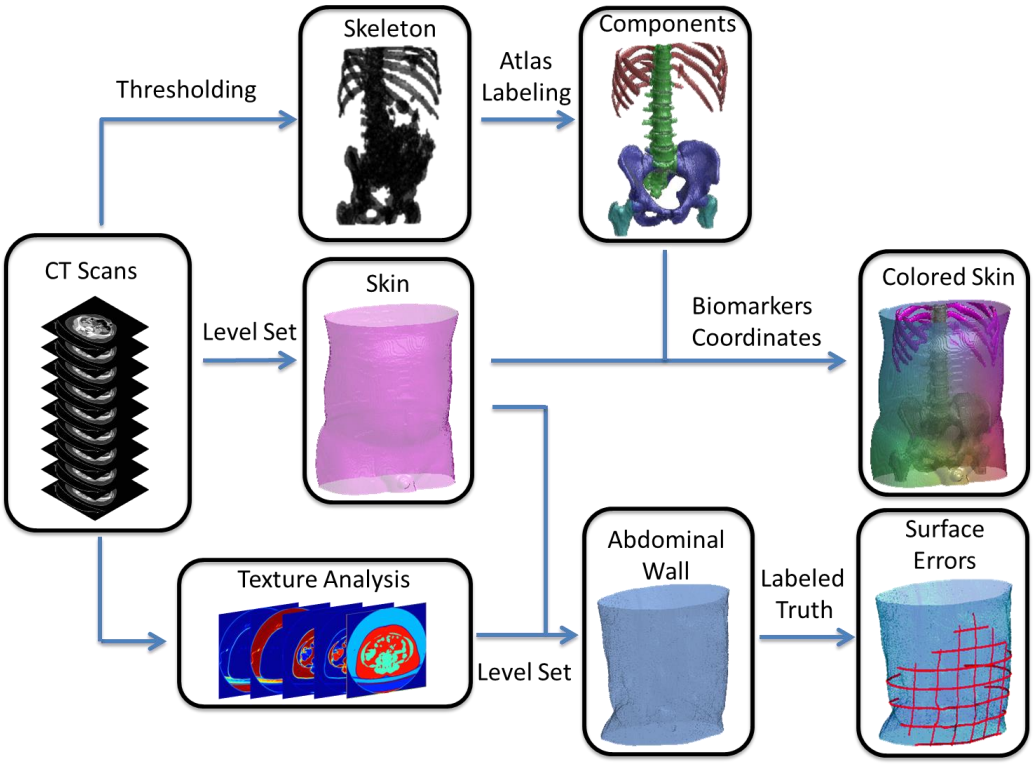

Figure V.2. Flowchart of the proposed method. The target image was affine-registered to a probabilistic atlas in terms of the extracted high intensity structures. Specific types of bony structures were then identified by transferring labels from the atlas to the target image based on a Bayesian framework, which incorporated the position information from the atlas and the intensity distribution for each label. After the skin is segmented by a curvature-constrained level set method, the two anterior iliac crests and all visible ribs are selected as landmarks, based on which three coordinates were created. The skin is then colored with RGB values converted by the normalized shortest distances to the biomarkers. Texture analysis followed by a fuzzy c-means procedure was used to estimate a voxel-wise probabilistic membership. An edge map was derived from the membership to guide the level set evolution, while the hard segmentation of muscles from the membership combined with the segmented skin was used to derive the initial start. Ground truth was manually labeled for the abdominal wall to calculate the surface errors of automatic segmentation.

intensities between muscles and adipose tissue in CT scans, an intuitive approach could be to extract the muscles by intensity thresholding, and then use a series of appropriate morphological operations to obtain

the whole abdomen region surrounded by the muscles. However, three facts can greatly challenge its reproducibility: (1) Skin can be difficult to eliminate due to variable thicknesses and variable distances to the muscles around the body across different subjects, given overlapping intensity distribution with muscles; (2) Muscles are greatly stretched in herniated regions and have much lower intensities than usual, which often results in gaps in muscle walls after thresholding; (3) Selection of morphological operators becomes challenging to construct a smooth abdominal region from thresholded images for the existence of imaging artifacts, especially in speckled adipose tissue.

Level set approaches, on the other hand, are more generally applicable with well-designed evolution functions. In the level set approach [116], an evolving surface is embedded as the zero level set of a higher dimensional level set function $\Phi(x,t)$, and propagates implicitly through the temporal evolution of Φ in terms of a given speed function F. $\Phi(x,t)$ is defined as signed distance function to the evolving surface with negative value inside the surface and positive outside. The speed function, which can be spatially varying, is usually determined by advection forces (e.g., constant inward or outward motion), intrinsic geometry (e.g., mean curvature), and image attributes (e.g., intensity, and its gradient). The temporal evolution of the level set function is usually described in the following form,

$$\Phi_t - F|\nabla\Phi| = 0 \tag{5.1}$$

where $|\nabla\Phi|$ represents the normalized gradient of the level set function.

Region-based level set methods [119-121], which rely on the global homogeneity of spatial localized features and properties, are not well-suited to our problem because the abdominal region contains not only muscles, but also visceral adipose tissue, bones, and organs with and/or without contrast. Edge-based approaches [50-52], which rely on the local differences in the image, seems a reasonable approach to start with, to extract a continuous smooth outer surface of the anterior abdominal wall, although challenges (**Figure V.1**) remain to be dealt with.

Here we construct an evolution function that follows the Geodesic Active Contours (GAC) model [51],

$$\Phi_t = \omega_{balloon}F_{balloon} + \omega_{curv}F_{curv} + \omega_{edge}F_{edge} + \omega_{reg}F_{reg} \tag{5.2}$$

$$F_{balloon} = g|\nabla\Phi| \tag{5.3}$$

$$F_{curv} = g\kappa|\nabla\Phi| \tag{5.4}$$

$$F_{edge} = \nabla\Phi \cdot \nabla g \tag{5.5}$$

$$F_{reg} = \nabla^2\Phi - \kappa \tag{5.6}$$

where g is usually called edge-stopping function, proposed by Malladi et al. [122] to stop the level set evolution at the boundaries of objects, which can be typically measured for an image I in the form as follows,

$$g = \frac{1}{1 + |\nabla I|^2} \tag{5.7}$$

κ represents the local curvature of the signed distance function

$$\kappa = \mathrm{div}\frac{\nabla\Phi}{|\nabla\Phi|} \tag{5.8}$$

In the evolution function shown in (2), $F_{balloon}$ acts as an inward/outward normal pressure force, also known as balloon force, to speed up the evolution process. F_{curv} is a curvature-constraint term to smooth the level set evolution adjusted by edge stopping function. F_{edge} is a term to reinforce the edge-preserving capability, especially under the balloon force. We append a regularization term F_{reg} proposed by Li et al. [175] to enable the level set evolution free from additional re-initialization. The weights associated with the four terms can be customized for specific purposes.

Local optimum and boundary leaking problem are the two biggest issues for edge-based level set methods. Particularly, the evolution of edge-based level set is likely to be stuck at noisy regions away from the target boundary, or on the contrary, pass through the target boundary where the contrast is not high enough. Both cases can be present in our VH data. This leaves the edge stopping function and the initial start of level set critical to the segmentation results.

The edge stopping function estimates the edginess across the image to guide the proceeding and stopping of the evolving surface. Regarding the image of edge stopping function as an edge map, we expect it to be clean and high-contrast, while the edge map derived from the original image can barely satisfy our expectation. Nonlinearly smoothing the image by an anisotropic filter, which averages each voxel with local voxels of similar intensity, can be an option to reduce the noise in an image while

preserving the edges between regions. Yao et al. [32] proposed to classify the body into adipose tissue and non-adipose tissue by fuzzy c-means (FCM) clustering on smoothed images, and derived the edge map by the probabilistic membership of the two clusters instead of intensity. However, these intensity-based efforts may be less effective when the intra-tissue variability is nearly on par with the inter-tissue variability, which is not uncommon in VH CT scans.

We focus on textures. Adipose tissue is embedded with scattered speckles and is also distinguishable from muscle groups constructed with muscle fibers, and solid bony structures. In addition, streaking artifacts are more severe in adipose tissue than in muscles and bones. Therefore, we consider texture analysis a potential tool for tissue classification around the abdominal region. The frequency- and orientation-selective properties of a Gabor filter provides us a multi-channel approach to extract texture features at multiple scales. A Gabor filter can be considered as the product of a Gaussian envelope and a sinusoid function [176]. For our specific use, we use a 3-D Gabor filter in the form of

$$h(x', y', z') = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma^3} e^{-\frac{1}{2\sigma^2}\left(x'^2+y'^2+z'^2\right)} \cdot \cos\left(\frac{2\pi x'}{\lambda} + \phi\right) \tag{5.9}$$

where, λ and ϕ are the wavelength and phase offset of the sinusoid function, respectively, σ is the standard deviation of the spatially isotropic Gaussian envelop, specified as a dependent variable of λ (i.e., $0.56\,\lambda$), and $(x', y', z')^T$ are the rotated spatial coordinates of Gaussian envelope,

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{5.10}$$

where θ denotes the orientation of the Gabor filter in xy plane. Note that we only take the in-plane texture features into consideration by the setting of sinusoid function in (9) and coordinates rotation in (10).

Multiple filtered images can be obtained by convolving the original image with a bank of Gabor filters. In practice, we set the phase offset as zero, choose $[0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}]$ as four possible values for orientations, and follows a frequency selection scheme proposed by Zhang et al. [177], which emphasizes on intermediate frequency bands. Suppose the size of the cross-sectional image is $N \times N$, the selected frequencies F can be estimated as follows,

$$\frac{1}{\lambda} \equiv F = 0.25 \pm {}^{2^{i-0.5}}/_{N}, \qquad i = 1, 2, \dots, \log_2({}^{N}/_{8}) \tag{5.11}$$

Filtered images are converted into feature images by a nonlinear transformation, $\psi(t) = \tanh(\alpha t)$ with α set as 1, where the sinusoidal modulations in the filtered images are transformed into square modulations, or in other words, the features are enhanced [60]. The stack of all these feature images forms a feature vector for each voxel of the original image, which enables us to use FCM clustering for classification. Here we empirically set the number of clusters as eight, and obtain the

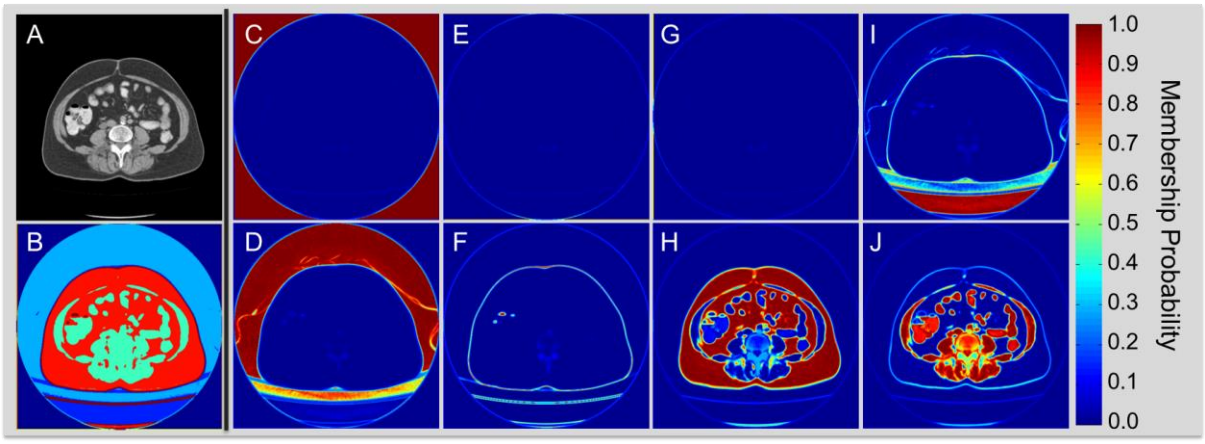

Figure V.3. Proposed texture analysis. (A) Original CT image; (B) Hard segmentation of different structures; (C) – (J) illustrate the membership probability for each of the eight clusters, which are estimated from fuzzy c-means clustering on texture features extracted by Gabor filters. Within each cluster, the probability value indicates partial membership to the cluster. Note that fat tissue and muscles can be identified from (H) and (J), respectively. (B) is constructed by the modes of among the eight clusters for all voxels, where we note that the muscles and fat tissue are effectively partitioned for assistance of the following edge-based level set segmentation.

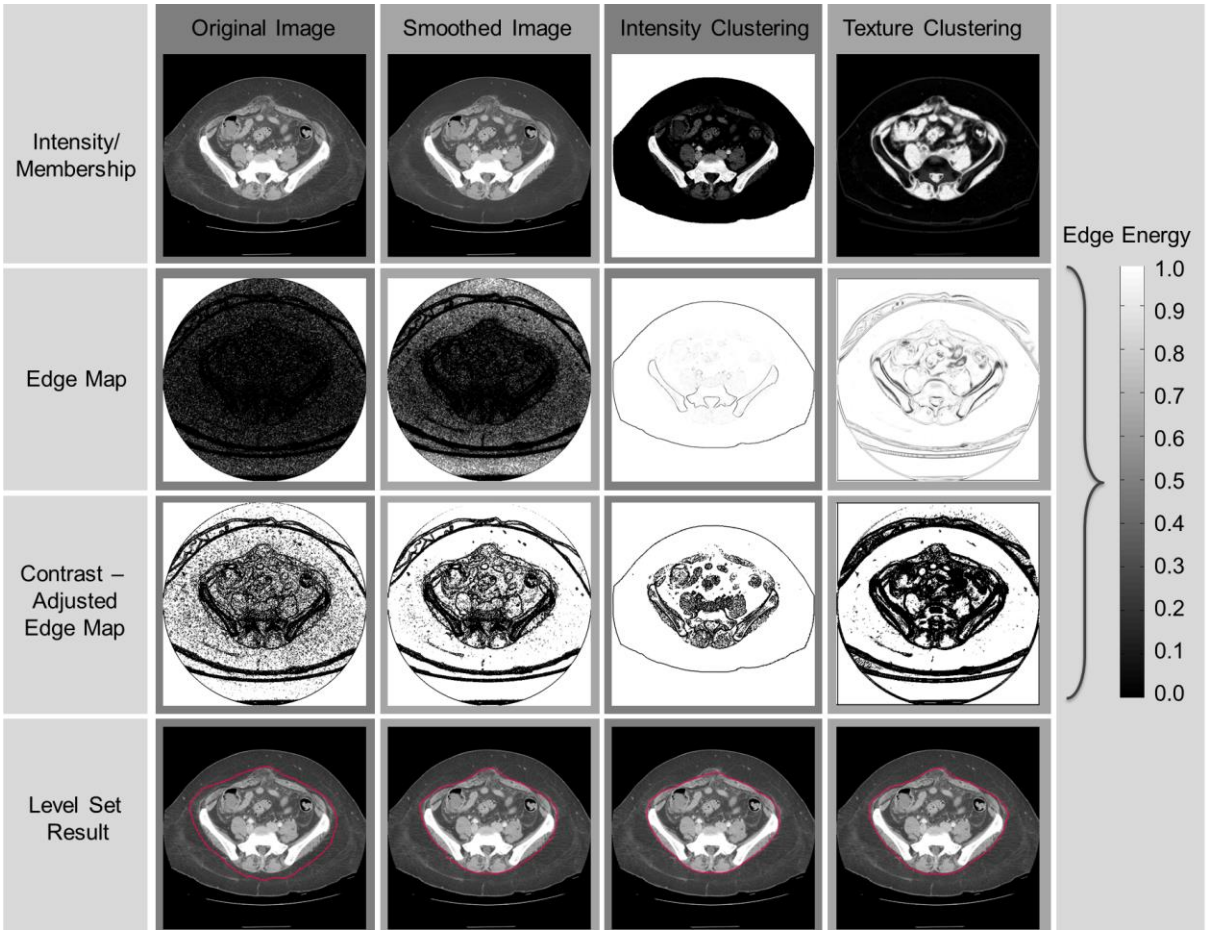

Figure V.4. Edge map and level set results of four methods. The four columns illustrate the results on the original image, the image smoothed by an anisotropic filter, the membership estimated by an intensity-based fuzzy c-means clustering, and the membership estimated by a texture-based fuzzy c-means clustering, respectively. The first row shows the intensity images (first two columns) and membership images (last two columns). The second row shows the edge maps directly derived from the images of the first row. The third row illustrates the contrast-adjusted edge maps for ease of level set evolution. Note that the level set front tends to proceed at brighter regions, and to stop at darker regions on the edge map. The fourth row presents the level set results (represented with red contours) on anterior abdominal wall segmentation for four methods with the same parameters.

probabilistic membership for these eight clusters as soft segmentation (**Figure V.3**). We define M_i as the probabilistic membership for the i^{th} cluster, and N_c as the number of clusters, and characterize the edge

map in terms of the texture,

$$g_t = \frac{1}{1 + \sum_{i=1}^{N_C} |\nabla \mathrm{M}_i|^2} \tag{5.12}$$

where $|\nabla \mathrm{M}_i|^2$ indicates the local difference based on each texture membership as opposed to intensity defined in (7) so that edges between distinct textures can be enhanced. Note that we sum over all probabilistic membership to ensure that all texture-wise differences are captured.

The contrast of the edge map is enhanced by an approximated Heaviside function for ease of edge-based level set, where the edge map is expected to be almost zero at abdominal walls, while approaching one for adipose tissue. Comparing to other three methods using (1) original image; (2) smoothed image; (3) intensity clustering, the edge map derived from the texture analysis provides better contrast, i.e., strengthened wall boundaries with reasonably cleared adipose, which enables the level set to evolve right at the abdominal wall, rather than get stuck by non-muscle structures or break into the abdominal cavity (**Figure V.4**).

We note that within the hard segmentation (**Figure V.3B**), i.e., the voxel-wise mode cluster member of the soft segmentation, muscles are quite well-distinguishable from skin in terms of texture clusters as opposed to the scenario of intensity thresholding, which leaves a good opportunity to derive an initial start of level set evolving surface close enough to the abdominal wall. In particular, we start with the largest component of the hard-segmented muscles (**Figure V.5, 1st col**), and filled the holes within the surface 30 voxels outside the zero level set surface of the muscle (**Figure V.5, 2nd col**). Then from the surface of the hole-filled volume, we go inside by 20 voxels, and eliminate the regions outside the segmented skin surface. The surface of the rest volume (**Figure V.5, 3rd col**) is considered as our initial start for the following level set evolution (**Figure V.5, 4th – 6th col**) for abdominal wall segmentation. Note that all the distances discussed here are in the unit of voxel in 3D.

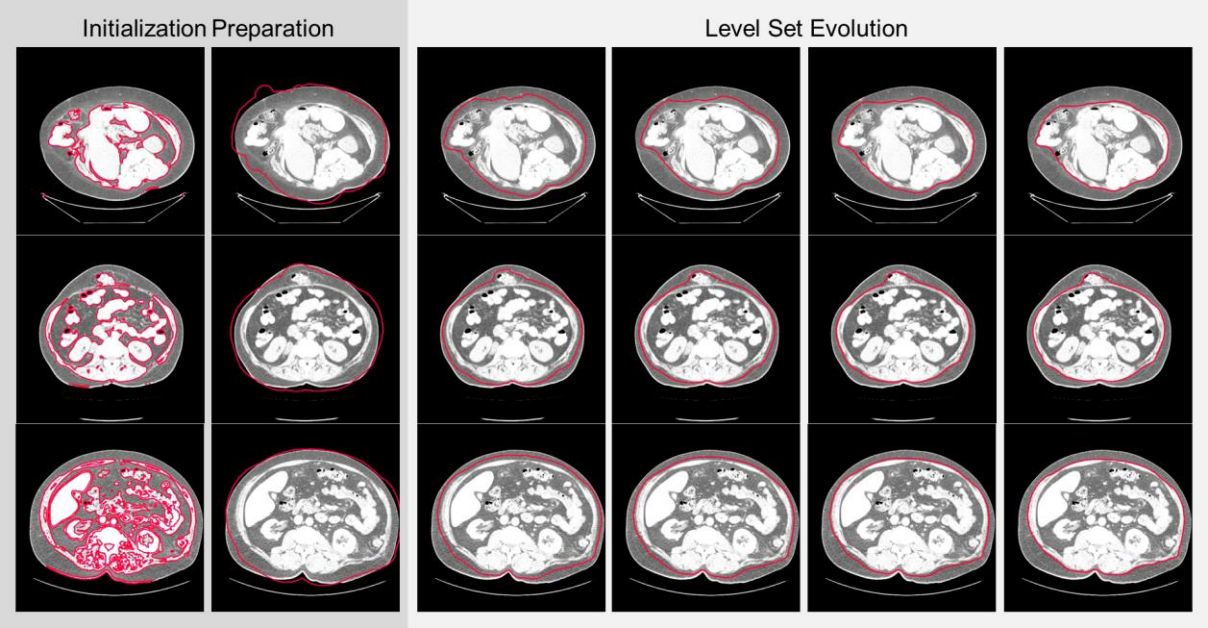

Figure V.5. Process of level set segmentation on anterior abdominal wall for three subjects. The red contours indicate the current segmentation for each process. The first two columns show the preparation for the initial start. The last four columns illustrate the iterations of level set evolution. Note the third column demonstrates the initial start of the level set segmentation.

2.2. Implementation

Detailed parameters during the implementation for the segmentation of abdominal wall not covered above are given as follows.

We enhance the contrast of the edge map from the proposed texture analysis by an approximated Heaviside function,

$$y = \frac{1}{2}\left[1 + \frac{2}{\pi} arctan\left(\frac{x - x_0}{\varepsilon}\right)\right] \tag{5.13}$$

where x_0 is the threshold, and ε is a small number which determines the steepness of the Heaviside function. For our method after texture clustering, we assign x_0 and ε to be 0.995 and 0.001, respectively.

Weighting parameters are specified as 5, 50, 5, and 10 for the four terms in (2), i.e., balloon force, curvature-constraint, edge-preserving, and regularization, respectively. We run the level set evolution on the whole volume for 500 iterations.

The other three methods we test for comparison share the same weighting parameters, iteration number, and the same initial start as our texture clustering method, which leaves the difference only on the edge maps. Specifically,

(1) Original image: We measure the edge map from the raw CT image, and enhance its contrast by applying (13) with $x_0 = 0.005$ and $\varepsilon = 0.001$.

(2) Smoothed image: We smooth the raw CT image with an anisotropic filter via SUSAN (Smallest Univalue Segment Assimilating Nucleus) [178] in FSL (FMRIB Software Library, University of Oxford, Oxford, UK). We specify the anisotropic filter a spatial size of 10 mm, and a brightness threshold of 20, under which is considered as noise to eliminate. The edge map measured from the smoothed image is enhanced by applying (13) with $x_0 = 0.005$ and $\varepsilon = 0.001$.

(3) Intensity clustering: We cluster the intensity of the smoothed image into adipose and non-adipose regions by FCM following the two-step procedure described in [32]. The edge map derived from the probabilistic membership of the two clusters is enhanced by applying (13) with $x_0 = 0.995$ and $\varepsilon = 0.001$.

2.3. Other Structures

Bony skeleton and skin surface are also essential structures for VH characterization; however, our approaches for the segmentation of these structures are not directly relevant to the texture-features in the level set segmentation. Therefore, only brief descriptions of our approaches are provided as follows.

We consider parts of the bony skeleton as natural structures to derive landmarks for VH localization [34]. Precise surface extraction of bones is beyond our major concern. Instead, we separate the whole volume into five partitions – (1) background; (2) pelvis; (3) spinal vertebrae; (4) femurs; (5) ribs, and any other bones surrounding the spine (e.g., sternum, scapula, etc.). We use an atlas-based segmentation method to partition the bony skeleton. Briefly, we first align patients' poses in CT scans with a pre-constructed atlas in terms of the high intensity (> 200 HUs) anatomical structures by affine

registration using FLIRT (FMRIB's Linear Image Registration Tool [179], University of Oxford, Oxford, UK). Then we combine the empirical intensity distribution of each structure (as the generic likelihood) with the position probability obtained from the smoothed atlas (as the prior) into a Bayesian framework so that the bony structures not perfectly aligned, but close to the atlas may still be captured by its inherent intensity, while non-bone structures hit by the registration can be cleared out.

Skin surface can be considered as a starting point for further imaging segmentation (i.e., the abdominal wall) inside human body, while skin segmentation, or body extraction in CT scans, can be reduced to the problem of removing the table and sheet after intensity thresholding since the majority of the human body (except for the air) in CT scans is of intensity larger than -250 HUs. Intensity clustering is a simple and efficient approach for body extraction [32]; however, there are cases that patients touch, or even lay right on the table without any sheet in between so that the table can be connected with the human body after intensity clustering, where we believe level set with curvature constraint is a more practical option. Briefly, we construct the evolution function with an intensity constraint (-250 HUs ~ 50 HUs) to guarantee the inclusion of the adipose tissue (usually -100 ~ -50 HUs) while excluding the background (around -1000 HUs), and a curvature constraint to smooth out the surface temporarily stuck at the table edge during the level set evolution. We run the level set evolution on every slice of the volume for 400 iterations, initializing with a box four pixels into the image boundary. After each iteration, we update the Φ values within a narrowband of 2 pixels around the zero level set, and recalculate the signed distance function based on the largest component of the regions with non-positive Φ values as an approximation of a re-initialization process.

Based on the segmented bony skeleton and skin surface, we calculate putative patient-specific coordinates on the purpose of a robust localization metric across subjects. We consider the left and right iliac crests (the most anterior point on each side of iliac spines) extracted from pelvis, in addition to all the rib bones as three groups of landmarks, to which we calculate the shortest spatial distances to construct three coordinates.

3. Methods and Results

Retrospective, clinically acquired CT data on 20 patients with suspected VHs were acquired in anonymous form under institutional review board supervision. The including criteria were to select patients having abdominal scans that covers from xiphoid process superiorly to pubic symphysis inferiorly. Large variations were seen among the volumes in voxels (512x512x90 ~ 512x512x200) and resolution (0.6x0.6x5 mm ~ 1.0x1.0x3 mm). Average field of view in millimeters was approximately 400x400x500 mm.

The ground truth of abdominal wall was created by a research associate, who was trained on a manual labeling protocol [34] for hernia-related anatomical structures using the Medical Image Processing And Visualization (MIPAV) [153] software (National Institutes of Health, Bethesda, MD) and a high resolution tablet input (Wacom, Tokyo, Japan) on a 64-bit Linux workstation. The entire herniated region was labeled volumetrically, while the normal abdominal wall was labeled on axial and sagittal slices approximately spaced every 5 cm for efficiency. We applied thin-plate spline interpolation to the label meshes of abdominal wall combined with the hernia label, to provide the surface of anterior abdominal wall, as well as the outer surface of herniated region and normal wall as two subsets for validation. We note that the choice of manual validation protocol was designed to yield an acceptable accuracy of anatomical labeling within one hour of manual time per patient. We have found that thin plate spline interpolation in the normal wall provide resolution on par (+/- 2mm) with manual tracing with a 5mm gap. The curvature within the herniated regions exhibits much higher spatial resolution, so every slice was labeled.

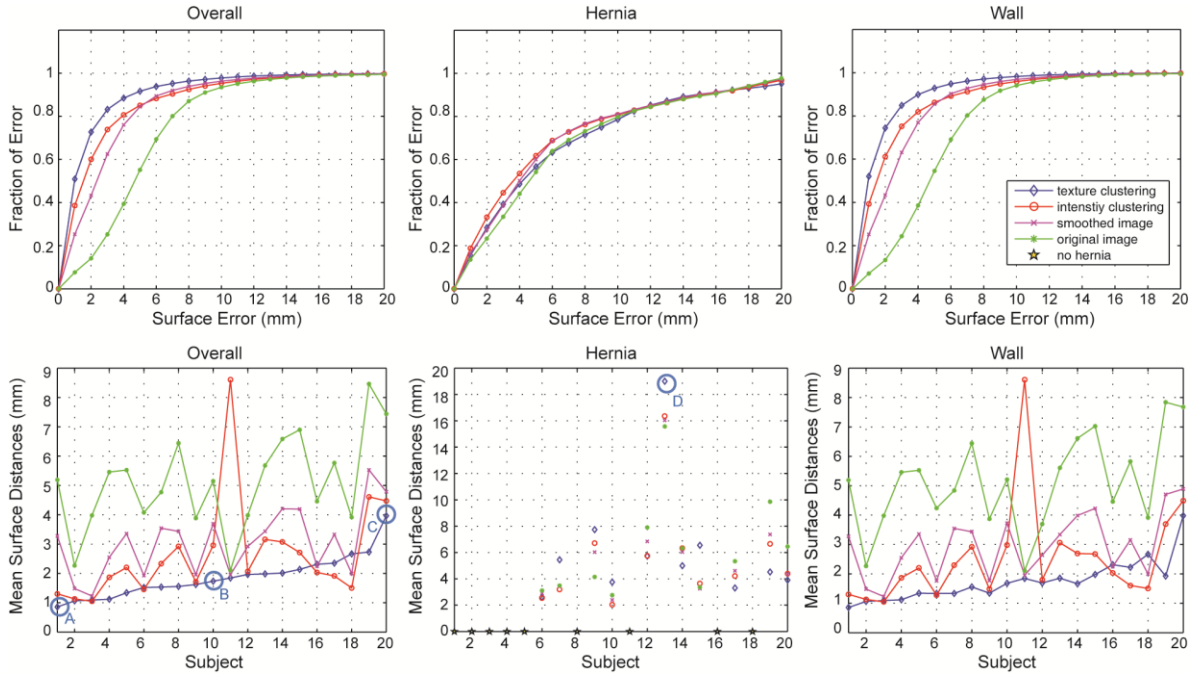

Figure V.6. Quantitative results of anterior abdominal wall segmentation. Four level set methods based on different edge maps (1. Green star: baseline original image; 2. Magenta cross: smoothing image; 3. Red circle: intensity clustering; 4. Blue diamond: texture clustering) are evaluated in terms of surfaces distances. The first column denotes the error metrics for the whole wall, while the second and the third column focuses on the hernia region and the normal wall region, respectively. The top row illustrates the cumulative fraction of region based on increasing 3D surface distance error between interpolated wall from manually labeled ground truth and the automatic segmentation. The bottom row shows the error bar of mean surface distances. The subject indices are sorted in terms of the mean of the overall mean surface distances of the texture clustering method. Note that 9 out of 20 subjects have no hernia labeled in the truth, which is indicated as gold stars in the middle-bottom plot. Four subjects (A, B, C, D) are selected to illustrate the qualitative results in Figure V.7.

Bony structures, skins, and anterior abdominal walls were extracted from the 20 selected scans with our proposed automated segmentation methods. Here the bony structures are regarded as visual references, while skin segmentations serve as the loose outer boundaries of abdominal walls. For abdominal wall segmentation, we integrated the edge map and the initial start derived from the texture

analysis into the baseline GAC level set model. For comparison, we tested three other methods using only intensity values using the same GAC model, with the same parameters in the level set evolution function, and for simplicity, starting from the same initial surface provided by our texture analysis. The segmentation results were validated against the ground truth, where surface distances were calculated from every point of the interpolated abdominal walls to the automatically segmented walls.

Using texture, median surface errors were ±1mm for the abdominal wall and <5mm for the hernia; errors were significantly greater (2-5 mm) for methods that did not use the texture (**Figure V.6, 1st row**). The mean surface errors across 20 subjects were 1.87±0.72 mm over the anterior abdominal wall using our texture-integrated method, which was statistically significantly lower ($\Delta = -0.78, p < 0.05$) than the best case of the three intensity-based methods in terms of paired Wilcoxon signed rank test (**Figure V.6, 2nd row**). **Figure V.6** and **Table V.1** provide more detailed error metrics on four tested methods.

Four subjects with variable mean surface errors were selected to illustrate the performance of the proposed method. Segmentation results were demonstrated in a volumetric view, as well as in slices, both with ground truth overlaid. Generally, the entire abdominal regions of the four subjects were well segmented with reasonable errors around the herniated regions (**Figure V.7**). In addition, we rendered the

Table V.1. Error metrics based on mean surface distance (MSD).

Method	Region	MSD (mm)	MSD < 1mm (%)	MSD < 2mm (%)	MSD < 5mm (%)
Original image	Overall	5.1±1.61	7.65±9.55	14.09±15.09	55.12±22.12
	Hernia	6.21±3.82	13.57±8.71	23.33±14.55	54.2±29.82
	Wall	5.09±1.58	7.1±9.79	13.29±15.48	54.56±22.96
Smoothed image	Overall	3.05±1.13	25.31±15.19	43.25±19.47	84.43±10.8
	Hernia	5.75±3.79	16.47±12.37	27.42±17.71	60±25.2
	Wall	2.96±1.07	25.18±15.5	43.39±19.99	85.43±10.03
Intensity clustering	Overall	2.65±1.71	38.66±16.29	60.05±19.13	85.07±18.22
	Hernia	5.62±3.92	18.81±10.36	33.2±16.92	61.78±26.01
	Wall	2.53±1.69	39.27±16.62	61.14±19.39	86.21±18.07
Texture clustering	Overall	1.87±0.72	50.89±14.5	72.7±13.71	91.61±4.96
	Hernia	6.14±4.52	15.52±10.5	28.36±16.21	56.65±25.79
	Wall	1.74±0.7	52.04±14.05	74.41±13.29	92.85±4.34

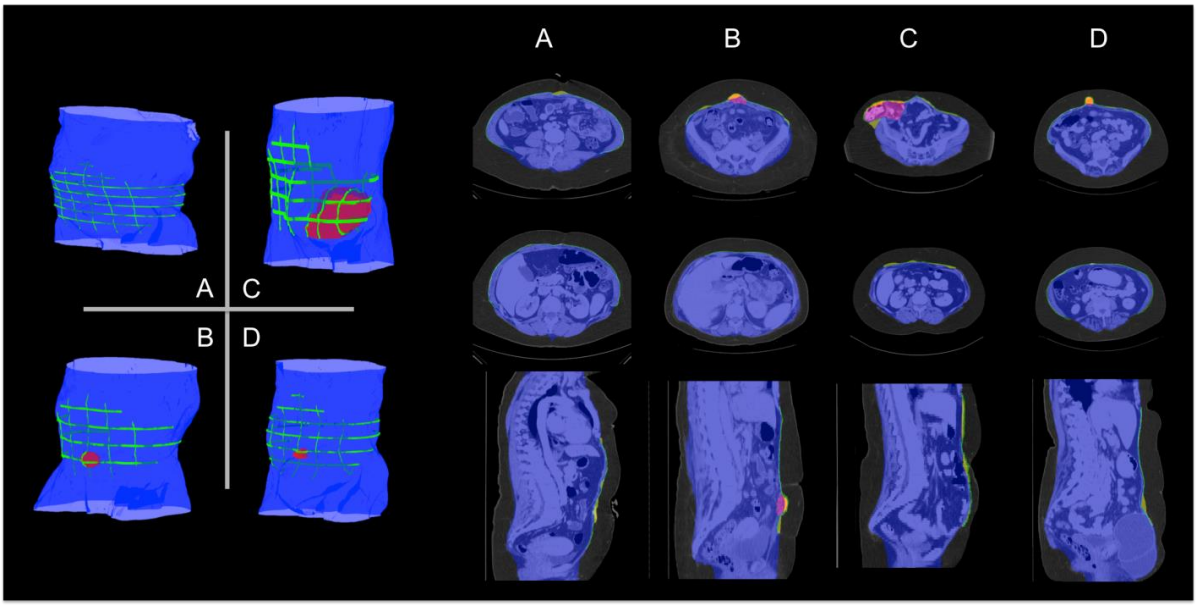

Figure V.7. Qualitative results on selected subjects from Figure V.6. The letters (A, B, C, D) match with the subjects circled in Figure V.6. The left panel shows a volumetric view of segmentation. The right part presents the results on several slices. Note blue denotes the automatic segmentation, green denotes the manually labeled anterior abdominal wall on sparsely sampled slices, and red denotes the manually labeled herniated region. In addition, in the slice representation, the segmentation errors for the normal abdominal wall are highlighted in yellow, while those for the herniated region are highlighted in orange.

interpolated abdominal wall with surface errors for the whole cohort. Over half of subjects had acceptable errors for the entire wall (**Figure V.8**).

4. Discussion

An edge-based level set method integrated with texture analysis is proposed to extract the anterior abdominal wall, which provides the segmentation with mean surface errors less than 2mm on 20 retrospective subjects validated by manually labeled ground truth. We do not claim that the proposed method has completely addressed the most challenging cases, but these are less impacted compared to other published methods. Specifically, based on the qualitative results shown in **Figures V.5, V.7, V.8**, we believe that challenges represented by **Figure V.1G-1I** are well handled by the proposed method,

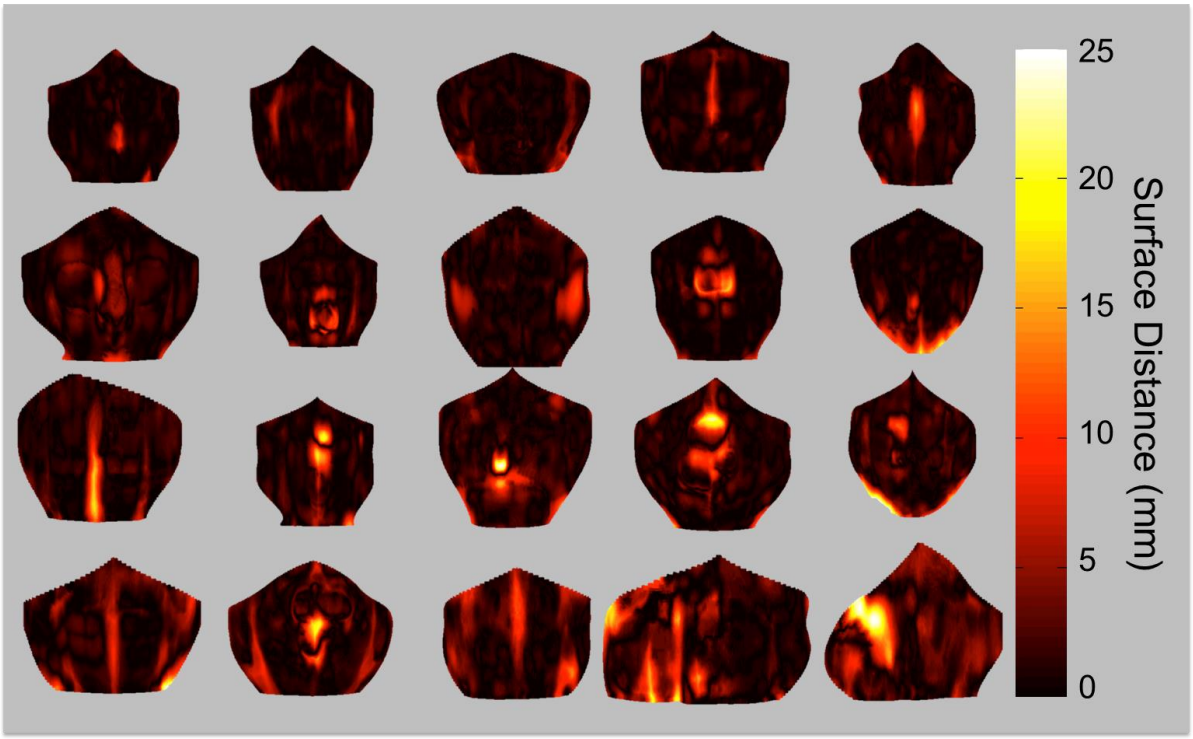

Figure V.8. Error maps for 20 subjects. The shape of the error map is provided by a thin –plate spline interpolation of the anterior abdominal wall on manual labeled meshes, where the lateral boundaries of the interpolated surface is also given by thin-plate spline interpolation, but on the terminations of the label meshes. The rendering color represents the 3D surface distance from the automated segmentation to the interpolated abdominal wall. Note that errors are most prominent in fascia and hernia regions.

while the challenges of **1C-1F** and **1J** are under better control than other methods. The methods using original image or smoothed image tend to suffer from the problem of **1G-1I**, while the method of intensity clustering is more likely to present worse performances in the cases of **1C-1E**. In addition, quantitative results in **Figure V.6** and **Table V.1** show the improvement of the proposed method on a global scale. Therefore, we come to a conclusion that the inherent texture patterns are helpful to the tissue classification, and texture analysis can improve the level set segmentation around the abdominal region.

The capability of identifying muscles from fat is critical to the proposed segmentation of the abdominal wall. Through the texture analysis, the majority of muscles that constructs the abdominal wall can be distinguishable from fat tissue. The fascia regions, as well as some herniated regions appear as

similar texture as the fat tissue (**Figure V.3H, 3J**). This creates gaps between the muscles of abdominal wall, which challenges the level set segmentation for not intruding into the abdomen. Curvature constraints in level set cost function serve to prevent the intrusion by smoothing the surface evolution, but may fail to capture the underlying structure (**Figure V.8**). The identification of those regions requires further study, which can promisingly improve our current segmentation of the abdominal wall. We also note that some internal organs, e.g., kidney and small bowel, have similar texture as fat. This might be an issue for the visceral fat; however, it is not critical to the segmentation of surface of the abdominal wall. Therefore, we are more focused on the quality of texture classification between the skin surface and the outer surface of abdominal wall, i.e., the region that our level set method evolves through, than the internal abdomen.

The segmentations of bony skeleton and skin present visually reasonable results. Robust landmark derivation from these structures is under development for VH characterization, and quantitative validation will be conducted on the derived landmarks. In continuing efforts, we are interested in characterizing the abdominal wall composition, including the tissues in the narrowband (small area) around the outer fascia surface (including muscle, bone, fascia, organ, air, etc.). Using surface curvature models and tissue classifications, we will seek to identify abnormalities and disruptions (hernias) in the abdominal wall. Ultimately, improved quantification of hernia and abdominal wall structure will lead to a better idea of whom to treat (or not treat) and with what type of repair. This would lead to standardization of care and ultimately, less waste in this commonly performed procedure that is currently rife with unnecessary variation in care.

CHAPTER VI

AUGMENTED ACTIVE SHAPE MODEL (AASM) COMBINING MULTI-ATLAS LABEL FUSION AND LEVEL SET

1. Introduction

Segmentation of human anatomical structures is challenging in medical images due to their physiological and pathological variations in shape and appearance, the complicated surrounding context, and the image artifacts. Active shape models (ASM [66]), also known as statistical shape models [110], provide a reasonable approach to characterize the variations in human anatomy, and thus have been widely used in the medical image community [22, 180, 181]. Given training datasets and their representation of the shape (usually landmarks), statistical models can be established for the structure of interest to characterize (1) the modes of its shape variations and (2) the local appearances around its shape boundary to drive the segmentation on other images.

However, as a model-based approach, ASM may present catastrophic segmentation failures if configured inappropriately. Since the shape updates of ASM focus on only local context, ASM segmentation can be sensitive to the model initialization, and/or fall into local minimum when given a large search range for the updates, and thus undermines its performance. This problem can get worse in a common segmentation procedure of clinical acquired medical scans, e.g., computed tomography (CT) and magnetic resonance (MR) images given their highly variable contexts. We posit that integrating traditional ASM with global optimization can improve its robustness to those challenging problems.

There have been efforts to combine ASM with level set (LS) techniques, where shapes are implicitly represented by signed distance function (SDF), and the statistical models build on SDF using principle component analysis (PCA) are used to regularize the LS evolution [113, 114, 182]. Region-

based LS builds it speed function on global information, where the Chan-Vese (CV) algorithm [183] is most commonly used as to evolve the SDF by minimizing the variance both inside and outside the zero level set. Tsai et al. [113] augmented the CV algorithm with two additional terms in the speed function that penalized the deviance from pre-trained shape model to segment the left ventricle and prostate on MR. Despite its success in shape constraint, this approach was not built for accurate structural segmentation since no local appearance searching from traditional ASM was deployed to capture the boundaries of structures. In addition, it is difficult to design a LS speed function to handle more variable contexts, e.g., secondary structures located within and around the structure of interest.

The multi-atlas label fusion (MALF) technique [124, 184] has become popular recently for its robustness. Given the capabilities of the state-of-the-art registration tools to roughly match two images regardless of the underlying contextual complexity, MALF leverages canonical atlases (training images associated with labeled masks) for target segmentation by image registration and statistical label fusion. MALF, by design, provides not only the hard (categorical) segmentation, but also soft (probabilistic) estimation. Xu et al. [39] integrated MALF with shape constraints to improve spleen segmentation on CT by probabilistic combination; however, it was sensitive to the alignment between the MALF estimate and the shape model.

Here we propose an augmented active shape model (AASM) by integrating MALF and LS into the traditional ASM framework. Briefly using the AASM approach, the landmark updates are optimized globally via a region-based LS evolution applied on the probability map generated from MALF. This augmentation effectively extends the searching range of correspondent landmarks while reducing sensitivity to the image contexts, and thus improves the robustness of the segmentation. In the following sections, we present our proposed algorithm, and validate its efficacy on a toy example and two different clinical datasets. This work is an extension of a previous SPIE conference paper [185].

2. Theory

2.1. Problem Definition

Consider a collection of R training datasets (also called atlases in the context of MALF) including the raw images $\boldsymbol{I} \in \mathbb{R}^{N\times R}$ and their associated labels $\boldsymbol{D} \in L^{N\times R}$, where N is the number of voxels in each image, and $\boldsymbol{L} = \{0, 1\}$ represents the label for the background and the structure of interest, respectively (only considering binary cases for simplicity). Based on each training label, the shape of the structure is characterized by a set of n landmarks. The landmark coordinates $(x_1, y_1), \dots (x_n, y_n)$ are collected in a shape vector for each training dataset as $\mathbf{x} = (x_1,\ y_1,\ \dots, x_n,\ y_n)^T$. Correspondences of these landmarks across all training datasets are required. For a target image $\hat{\boldsymbol{I}}$, the goal is to provide a set of landmarks $\hat{\mathbf{x}}$ that represents the shape of the estimated segmentation $\widehat{\boldsymbol{D}}$.

2.2. Active Shape Model and Shape Regularization

The mean $\overline{\mathbf{x}}$, covariance $\mathbf{S}$ of the shape vectors of the training datasets are computed, where

$$\overline{\mathbf{x}} = \frac{1}{R}\sum_{i=1}^{R}\mathbf{x}_i\,, \mathbf{S} = \frac{1}{R-1}\sum_{i=1}^{R}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T\,. \tag{6.1}$$

Using PCA, the Eigenvectors $\mathbf{p}_i$ with its associated Eignenvalues λ_i are collected. Typically, Eigenvectors correspondent to the t largest Eigenvalues were retained to keep a proportion f_v of the total variance such that $\sum_{i=1}^{t}\lambda_i \geq f_v \sum_t \lambda_i$, where $\mathbf{P} = (\mathbf{p}_1|\mathbf{p}_2|\dots|\mathbf{p}_t)$. Within this Eigensystem, any set of landmarks can be approximated (often called shape projection) by

$$\mathbf{x} \approx \overline{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{6.2}$$

where $\mathbf{b}$ is a t dimensional vector given by

$$\mathbf{b} = \mathbf{P}^T(\mathbf{x} - \overline{\mathbf{x}}) \tag{6.3}$$

$\mathbf{b}$ can be considered as shape model parameters, and its values are usually constrained within the range of $\pm m_b\sqrt{\lambda_i}$ when fitting the model to a set of landmarks so that the fitted shape is regularized by the model.

2.3. Local Appearance Model and Active Shape Search

The intensity profiles along the normal directions of each landmark are collected to build a local appearance model to suggest the locations of landmark updates when fitting the model to an image structure. For each landmark in the i^{th} training image, a profile of $2k+1$ pixels is sampled with k samples on each side of the landmark. Following [186], the profile is collected as the first derivative of the intensity, and normalized by the sum of absolute values along the profile, indicated as $\mathbf{g}_i$. Assuming multivariate Gaussian distribution of the profiles among all training data, a statistical model is built for each landmark,

$$f(\mathbf{g}) = (\mathbf{g} - \overline{\mathbf{g}})\, \mathbf{S}_{\mathrm{g}}^{-1}\, (\mathbf{g} - \overline{\mathbf{g}}) \tag{6.4}$$

where $\overline{\mathbf{g}}$ and $\mathbf{S}_{\mathrm{g}}$ represent the mean and covariance, respectively. This is also called the Mahalanobis distance that measures the fitness of a newly sampled profile $\mathbf{g}$ to the model. Given a search range of m pixels ($m > k$) on each side of the landmark along the normal direction, the best match is considered with the minimum $f(\mathbf{g})$ value among $2(m-k)+1$ possible positions.

2.4. Multi-Atlas Label Fusion and Probability Map Generation

A pair-wise image registration is performed between each atlas $\boldsymbol{I_j}$ and the target image $\hat{\boldsymbol{I}}$ to generate a transformation $\boldsymbol{T}_{reg}^{j}$ that maximizes a similarity metric SM between the two images.

$$\boldsymbol{T}_{reg}^{j} = \underset{\boldsymbol{T}}{\operatorname{argmin}}\, SM\big(\boldsymbol{T}(\boldsymbol{I_j}), \hat{\boldsymbol{I}}\big) \tag{6.5}$$

This transformation is propagated on both the atlas image and label, where

$$\boldsymbol{I_j'} = \boldsymbol{T}_{reg}^{j}(\boldsymbol{I_j}), \boldsymbol{D_j'} = \boldsymbol{T}_{reg}^{j}(\boldsymbol{D_j}). \tag{6.6}$$

A label fusion procedure LF is then used on the registered atlases to generate a label-wise probabilistic estimation $\boldsymbol{W} \in \mathbb{R}^{N\times 2}$ on the target,

$$\boldsymbol{W} = LF\big(\hat{\boldsymbol{I}}, \boldsymbol{I'}, \boldsymbol{D'}\big) \tag{6.7}$$

where the registered labels $\boldsymbol{D}'$ are combined on a voxel (or pixel) basis, and typically weighted by the similarities between the registered images $\boldsymbol{I}'$ and the target image. The probability map of the structure of interest $\boldsymbol{M} \in \mathbb{R}^{N\times 1}$ can then be derived by normalizing $\boldsymbol{W}$

$$\boldsymbol{M} = \frac{\boldsymbol{W}_1}{\boldsymbol{W}_0 + \boldsymbol{W}_1} \tag{6.8}$$

where $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ represent the background and foreground probability respectively. We note that we leave some abstract notions (e.g., $\boldsymbol{T}_{reg}$, SM, and LF) in the description of MALF above given its sophisticated process and the large number of variants in implementation.

2.5. Level Set Evolution with Chan-Vese Algorithm

In the LS context, the evolving surface is represented as the zero level set of a higher dimensional function $\phi(x,t)$, and propagates implicitly through its temporal evolution (speed function) $\frac{\partial \phi}{\partial t}$ with a time step dt. $\phi(x,t)$ is defined as SDF, with negative/positive values inside/outside the evolving surface, respectively. The CV algorithm evolves the SDF by minimizing the variances of the underlying image u_0 both inside and outside the evolving surface. Given $C_1 = \text{average}(u_0)$ in $\{\, \phi \geq 0 \,\}$, and $C_2 = \text{average}(u_0)$ in $\{\, \phi < 0 \,\}$, the temporal evolution of CV can be written as

$$\frac{\partial \phi}{\partial t} = \alpha\delta(\phi)[\mu\kappa - (u_0 - C_1)^2 + (u_0 - C_2)^2] \tag{6.9}$$

where $\delta(\cdot)$ is the Dirac delta function, $\kappa = \text{div}\frac{\nabla\phi}{|\nabla\phi|}$ represents the curvature of SDF, α and μ are considered as the evolution coefficient and smoothness factor, respectively.

2.6. Augmented Active Shape Search

Given (1) trained active shape model, (2) trained local appearance model, and (3) probability map generated from MALF, an augmented active shape search procedure is performed in each iteration of the shape updates (**Figure VI.1**).

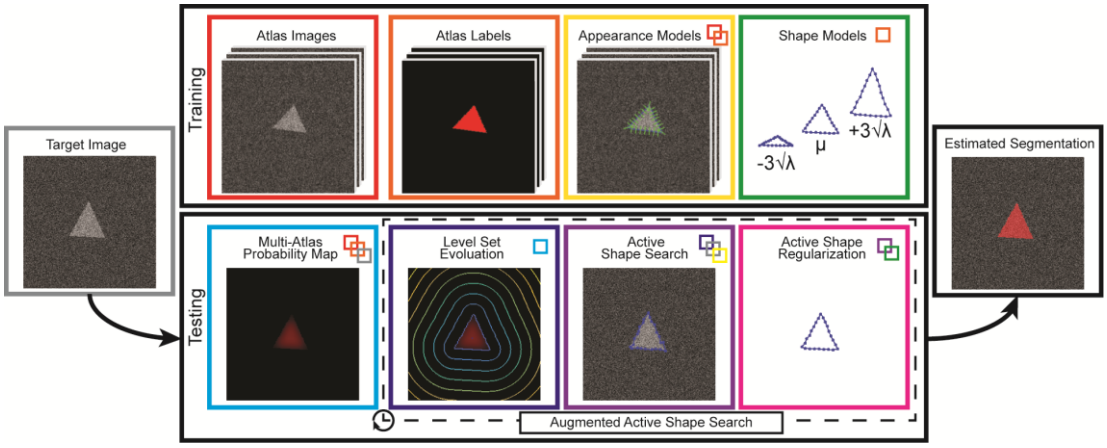

Figure VI.1. Flowchart of the proposed AASM approach. Shape models and local appearance models are constructed based on the atlas images and labels during the training stage. When testing on a target image to yield an estimated segmentation, an iterative process is performed. During each iteration, region-based LS is used to evolve on the probabilistic map generated by MALF to augment the traditional active shape search by global optimization, followed by the active shape regularization on the segmentation. Note that the borders of blocks are colored in distinctive colors. The small colored boxes within a block represent its prerequisite blocks in corresponding colors. For example, multi-atlas probability map requires (1) atlas images, (2) atlas labels, and (3) target image.

Let (x, y) be the current landmark position, ϕ_0 the current zero level set. First, LS evolution using Eq. 9 is performed by assigning $u_0 = \boldsymbol{M}$ for n_s iterations, and the zero level set moves to $\phi_0{}'$. Then the zero-crossing point along the normal direction of (x, y) on $\phi_0{}'$ is collected as (x', y'), and considered as the new landmark position after LS evolution. Along (x', y'), the gradient intensity profiles are sampled, then the active shape search suggests an updated position at (x'', y'') with its correspondent profile $\mathbf{g}'' = \underset{\mathbf{g}}{\operatorname{argmin}} f(\mathbf{g})$. The newly searched positions for all landmarks are then projected to the model space by Eq. 3. After constraining each model parameter, where $\tilde{b}_i = \min\left(\max\left(b_i, -m_b\sqrt{\lambda_i}\right), m_b\sqrt{\lambda_i}\right)$, the landmark positions are then regularized by Eq. 2. Note that n_s is a small number since the LS evolution here is used to suggest a globally optimal landmark movement as opposed to provide the final segmentation.

2.7. Optional Variants

The baseline pipeline can be optionally modified as follows to tailor for specific applications.

A multi-level scheme can be applied to the local appearance model, the LS evolution, and the active shape search. Typical downsample ratio is 2^{r-1} for the r[th] level; the landmark updates performed on the this level are with a multiplication of 2^{r-1} in the original image size, which effectively enlarges the search range.

An optional mask *B* can be used so that the evolution will only affected by the masked ROI, where the computation of the averages are modified as $C_1 = \text{average}(u_0)$ in $\{\phi \geq 0 \text{ and } B > 0\}$, and $C_2 = \text{average}(u_0)$ in $\{\phi < 0 \text{ and } B > 0\}$.

The landmark positions can be normalized by a transformation T_{norm} before deriving landmark positions from the evolved LS.

$$T_{norm} = \begin{bmatrix} r_{xs}/r_{xt} & r_{ys}/r_{yt} & c_{xs} - c_{xt} \\ r_{xs}/r_{xt} & r_{ys}/r_{yt} & c_{ys} - c_{yt} \\ 0 & 0 & 1 \end{bmatrix} \tag{6.10}$$

where (c_{xs}, c_{ys}) and (r_{xs}, r_{ys}) represent the centroid and the range along each dimension of the region within the zero level set, (c_{xt}, c_{yt}) and (r_{xt}, r_{yt}) represent the correspondent measurements for the current landmarks. With normalization, the landmark derivation from LS can be more robust to large shape updates.

Non-zero level set ϕ_v can also be considered as the surface to collect the updated landmarks after LS evolution to adjust the desirable intermediate segmentation in terms of $\boldsymbol{M}$.

3. Methods and Results

3.1. Toy Example

We defined a simulated observation consisting of three triangles in small, medium, and large sizes on a 256 x 256 image. We used the medium-sized triangle as the target of interest sandwiched by the other two triangles to increase its segmentation difficulty (see **Figure VI.2a**). An equilateral triangle

centered in the image was created as a shape template, where the radius of its circumscribed circle was 20 voxels. A 2-D affine transformation model T_{affine} with a rotational (θ), two translational (tx, ty), and two scaling components (sx, sy) was constructed as

$$T_{affine} = \begin{bmatrix} sx\cos\theta & -sy\sin\theta & tx \\ sx\sin\theta & sy\cos\theta & ty \\ 0 & 0 & 1 \end{bmatrix}. \quad (6.11)$$

All components were drawn from Gaussian distributions to generate randomized observations. Specifically, $\theta \sim \mathcal{N}(0,5)$ degrees, $tx \sim \mathcal{N}(0,2)$ voxels, $ty \sim \mathcal{N}(0,2)$ voxels, $sx \sim \mathcal{N}(s_b, 0.3)$, and $sy \sim \mathcal{N}(s_b, 0.3)$, where s_b was assigned as 1, 2, and 3 for the small, medium, and large triangles, respectively, as the base scale. The voxel-wise intensities over the small and large triangles were drawn from $\mathcal{N}(40,10)$, while those over the background and the medium triangle were from $\mathcal{N}(20,10)$. The datasets of our toy example include 100 randomly generated observations.

A leave-one-out cross-validation (LOOCV) scheme was used to validate the segmentation results of ASM and AASM. For each target observation, 99 other observations were used as training datasets. For each training data, 33 landmarks were evenly sampled on each side of the triangle (99 in total around the triangle). The active shape model was trained with 98% of the total variances, while the local appearance model was trained at 2 levels with an intensity gradient profile of 3 pixels collected along

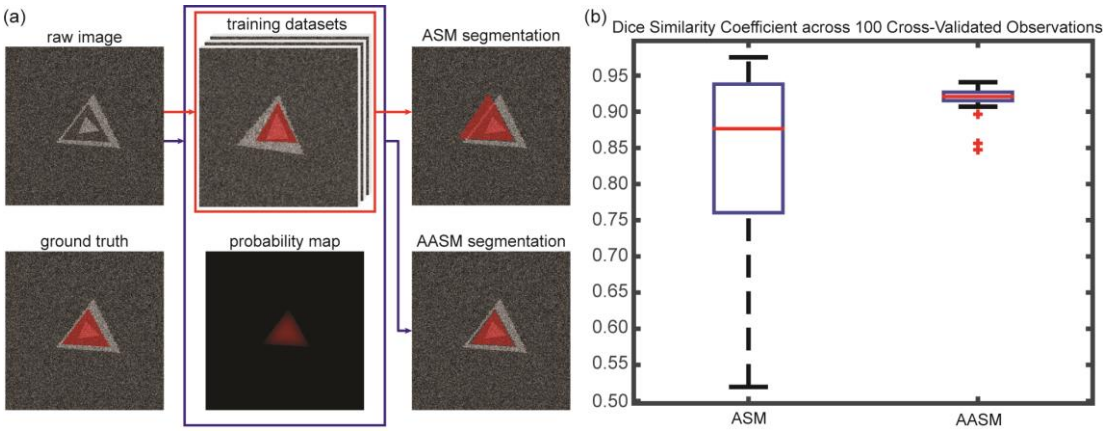

Figure VI.2. Results of a toy example. (a) Qualitative comparison between ASM and AASM segmentation on an individual observation. (b) Quantitative comparison between ASM and AASM segmentation in DSC across 100 cross-validated observations.

each side the normal directions of each landmark (7 pixels in total) at each level. During testing, both ASM and AASM were initialized by the mean model shape at the center of the image. In each iteration of the landmark update, the local search range was 6 pixels along each side of the normal direction. The shape updates were regularized within ± 3 standard derivations of the Eigenvalues over 100 iterations at 2 levels. For enabling AASM, a probability map was simulated by smoothing the ground truth by applying a 21 x 21 Gaussian kernel with a standard deviation of 5 voxels. 5 iterations of LS evolution were performed based on the simulated probability map with the time step, evolution coefficient, and smoothness factor setting as 0.01, 100000, and 0.00001, respectively, during each iteration of landmark update. The landmark positions were normalized based on the region within zero level set before determining the landmark movement based on the LS evolution.

Given the global optimization from the probability map, AASM was able to capture the correct boundary of the target shape, i.e., the medium triangle. Outlier segmentations in ASM were corrected in AASM based the DSC performances across 100 observations (**Figure VI.2b**).

3.2. Abdominal Wall

3.2.1. Data

Under institutional review board (IRB) supervision, abdominal CT data on 250 cancer patients were acquired clinically in anonymous form. These patients represent part of an overall effort to evaluate abdominal wall hernia disease in the cancer resection population. 40 patients were randomly selected, where we used 20 as training datasets, and the other 20 for testing purposes. The field of views (FOV) of the selected 40 scans range from 335 x 335 x 390 mm^3 to 500 x 500 x 708 mm^3, with various resolutions (0.98 x 0.98 x 5 mm^3 ~ 0.65 x 0.65 x 2.5 mm^3). Various numbers (78 ~ 236) of axial slices with same in-plane dimension (512 x 512) were found.

All 40 scans were labeled using the Medical Image Processing And Visualization (MIPAV [153]) software by an experienced undergraduate based on our previously published labeling protocol [34]. Following [35], essential biomarkers, i.e., xiphoid process (XP), pubic symphysis (PS), and umbilicus

(UB), were identified, and the abdominal walls were delineated on axial slices spaced every 5 cm with some amendments (contour closure required here). 177 and 184 axial slices were obtained with whole abdominal wall labeled for the training and testing datasets, respectively. Here we characterize the whole abdominal wall structure as enclosed by the outer and inner surface, bounded by XP and PS [35] This definition covers thoracic, abdominal, and pelvic regions, and includes not only the musculature, but also the kidneys, aorta, inferior vena cava, lungs, and some related bony structures to make the inner and outer boundaries anatomically reasonable.

3.2.2. Pre-processing

Given the large variations of appearance in abdominal wall and its surrounding anatomical structures along the cranial-caudal direction, the proposed slice-wise segmentation was trained and tested on five exclusive classes given the position of the axial slices with respect to XP, PS, and UB. These three biomarkers were acquired from manual labeling for the training sets, while estimated using random forest for the testing sets. We used 10 random scans from the training data to characterize the centroid coordinates of the biomarkers with long-range feature boxes following [187], and yielded the estimated biomarkers positions on the testing data with a mean distance error of 14.43 mm. Four bounding positions were empirically defined among the vertical position of the three biomarkers to evenly distribute the available training data (25, 35, 50, 31, 36 slices for each class, ordered from bottom to top). Given a target testing volume, each axial slice between the estimated positions of XP and PS was extracted, and assigned a class based on the estimated bounding positions. In this experiment, we only tested on the 184 slices with manual labels.

All slices (training and testing) were centered in the image after body extraction and background removal to reduce variations. A body mask can be obtained by separating the background with *k*-means clustering, and then filling holes in the largest remaining connected component. A margin of 50 pixels was padded to each side of the slices in case that the body was in contact with the original slice boundary, which made the slice size 612 x 612.

3.2.3. Training

On each training slice, landmarks were collected along the outer and inner wall contours using marching squares. The horizontal and vertical middle lines of the slice were used to divide each closed contour into four consistent segments across all slices assuming all patients were facing toward the same direction in the scan. 53 correspondent landmarks were then acquired on each of the segments via linear interpolation (212 for each of outer and inner wall). Each set of the landmarks was first centered to the origin, and then sets of landmarks from the same class were used to construct one active shape model covering 98% of the total variances. The local appearance model was trained at 3 levels; at each level, an intensity gradient profile of 5 pixels was collected along each side the normal directions of each landmark (11 pixels in total).

3.2.4. Testing

For each testing slice, all training slices from the same class were considered as atlases, and non-rigidly registered to it using NiftyReg [188]. The registered atlases were combined by joint label fusion

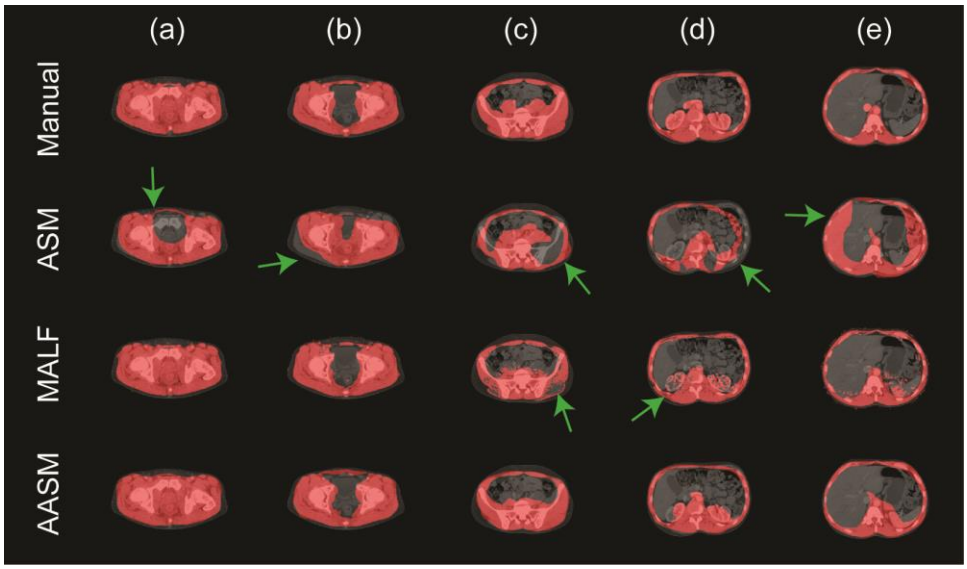

Figure VI.3. Qualitative comparison of ASM, MALF, and AASM segmentation of abdominal wall. (a) - (e) demonstrate slices in five exclusive classes on one subject. The green arrows indicate segmentation outliers including speckles, holes, over-segmentation, and label leaking problems.

Table VI.1. Abdominal wall segmentation metrics.

Method	DSC	MSD (mm)	HD (mm)
ASM	0.74 ± 0.15	9.70 ± 6.56	48.84 ± 17.52
MALF	0.89 ± 0.07	4.63 ± 2.40	46.03 ± 14.41
AASM	0.86 ± 0.09	4.77 ± 2.69	33.8 ± 15.09

(JLF) [134] to yield an probabilistic estimation of the abdominal wall. Default parameters were used for both. Within each iteration of landmark update, a region-based LS evolution with 5 iterations using CV algorithm was used to drive the landmark movement based on the global probabilistic estimation. The time step, evolution coefficient, and smoothness factor and was set to 0.01, 100000, and 0.1, respectively. The local search range for the landmark update was 8 pixels along each side of the normal direction. The shape updates were regularized within ± 3 standard derivations of the Eigenvalues. We allowed 100 iterations for 3 levels of shape updates.

3.2.5. Customized Configuration

In this study, a two-phase scheme was used to improve the robustness of whole abdominal wall segmentation. The proposed approach was first applied to only the outer wall. Initialized by the position of the outer wall segmentation, our approach was then applied to the combination of the outer and inner wall, while the outer wall landmark positions were fixed during the second phase shape updates. Active shape model and local appearance model were thus trained on (1) outer wall, and (2) outer and inner wall. The level set evolution for the second phase only considered the region within the outer wall segmentation obtained in the first phase.

3.2.6. Fat Measurement

Following [32], the fat tissue was obtained by using a two-stage fuzzy-c means. For each slice, the subcutaneous fat was considered as outside the outer surface of the abdominal wall, while the visceral fat as inside the inner surface.

Table VI.2. Abdominal Fat measurement errors.

Method	D_s (cm^3)	D_v (cm^3)	ρ_s	ρ_v	ρ_r	R^2_s	R^2_v	R^2_r
ASM	28.02 ±	36.52 ±	0.92	0.80	0.69	0.85	0.64	0.47
AASM	16.72 ±	15.38 ±	0.94	0.96	0.87	0.88	0.93	0.76

Note that D indicates the absolute difference between the area derived from the manual label and the estimated segmentation, ρ represents the Pearson's correlation coefficient, and R^2 is the R-square value of a linear regression. The subscripts s, v, and r represent subcutaneous fat, visceral fat, and the ratio of visceral fat to subcutaneous fat, respectively.

3.2.7. Results

The segmentation results were validated against the manual labels on 184 testing slices using Dice similarity coefficient (DSC), mean surface distance (MSD), and Hausdorff distance (HD) with comparison to results using ASM and MALF individually. Qualitatively, AASM presented the most robust result, while MALF has speckles and holes in the segmentations, or leak into the abdominal cavity where structures with similar intensities to muscles were present. ASM was sensitive to initialization, and could be trapped into local minimum (**Figure VI.3**). Quantitatively, large decreases in HD were observed when using AASM without undermining the DSC performance (**Table VI.1**). More importantly, the nature of ASM kept the topology of the abdominal wall, and enabled the compartmental fat measurement; MALF failed to do so even though it presented the best DSC performance. The absolute differences in subcutaneous and visceral measure using our augmented ASM against the measurement using manual labels were largely reduced comparing to traditional ASM (**Table VI.2**). In terms of the Pearson's correlation coefficient and R-Squared value between the estimated and manual measurement of subcutaneous fat, visceral fat, and the ratio of visceral to subcutaneous fat, AASM demonstrated consistent superiority over ASM (**Table VI.2**).

3.3. Spinal Cord

3.3.1. Data

With IRB approval, two batches of MR volumes of cervical spinal cord (SC) were acquired as training and testing datasets. The training datasets consisted of 67 scans of healthy controls, each

approximately covering a FOV of 224 x 224 x 90 mm^3 with 30 axial slices at a nominal resolution of 0.44 x 0.44 x 3 mm^3. The testing datasets included 28 scans, each approximately covering a FOV of 150 x 150 x 65 mm^3 with 10 ~ 14 axial slices at a nominal resolution of 0.65 x 0.65 x 5 mm^3, reconstructed to an in-plane resolution of 0.29 x 0.29 mm^2. Both datasets were acquired axially on a 3T Philips Achieva scanner (Philips Medical Systems, Best, The Netherlands), where T2*-weighted volumes were obtained using a high-resolution multi-echo gradient echo (mFFE) sequence to provide good contrast between the white matter (WM) and gray matter (GM) [189]. All scans generally covered the region from the 2^{nd} to 5^{th} cervical vertebrae, and the center of the image volume was aligned to the space between the 3^{rd} and 4^{th} cervical vertebrae. The in-plane dimensions for axial slices ranged from 512 x 512 to 576 x 576. 17 of the 28 testing subjects were diagnosed with multiple sclerosis (MS). Local and diffuse inflammatory lesions were presented for MS patients throughout the cervical cord white matter with similar appearance to GM in mFFE images [190].

The ‘‘gold standard’’ manual labels were constructed on both datasets. For each slice, two labels, i.e., WM and GM., were considered: The labeling process was performed using MIPAV [153] by an experienced rater on a collection of 1538 slices with reasonable contrast (not all slices due to image artifacts) in the training datasets, while using FSLView [191] by another experienced rater on all 364 slices in the testing datasets; both raters were familiar with MR images of the cervical SC.

3.3.2. Pre-processing

Here we integrated AASM into the slice-wise spinal cord segmentation framework in [192] with some essential modification.

A common region of interest (ROI) with the size of 91x77 was created using the 1538 co-registered training slices given the extent of the manual spinal cord labels. All testing slices were transferred into this space before the segmentation.

A volume-wise initialization was first performed using 2-D convolution. Consider the averaged image of the testing volume along the cranial-caudal axis as A, the average image of all cropped training

image slices as B, a 5 x 5 matrix with all entry values as 1/25, the highest response point of $A * B * C$ was identified as the approximate centroid of spinal cord for the testing volume, and considered as the starting point for the following slice-wise registration.

The slice-wise registration followed [192]. Briefly, an active appearance model was created using the cropped training images to capture the modes of variations of the spinal cord appearance within the ROI. A target slice was projected to the low-dimensional model space. It was then registered using a model-specific cost function given the differences in intensities and model parameters between the current estimate and the closest cropped training images. The registration searched at 5 levels (coarse to fine) over the three degrees of freedom (DoF), i.e., two translational (± 30 and ± 80 mm along x and y axes, respectively) and one rotational component ($\pm 10°$). At each level, the registration was optimized using line search on each DoF, followed by Nelder-Mead simplex method on all DoFs.

After registration, 30 cropped training images closest to the target image in the model space were selected. These images were used as the target-specific training sets for building ASM, as well as the atlases for MALF in the following process. Note given our proposed surface-based approach, we

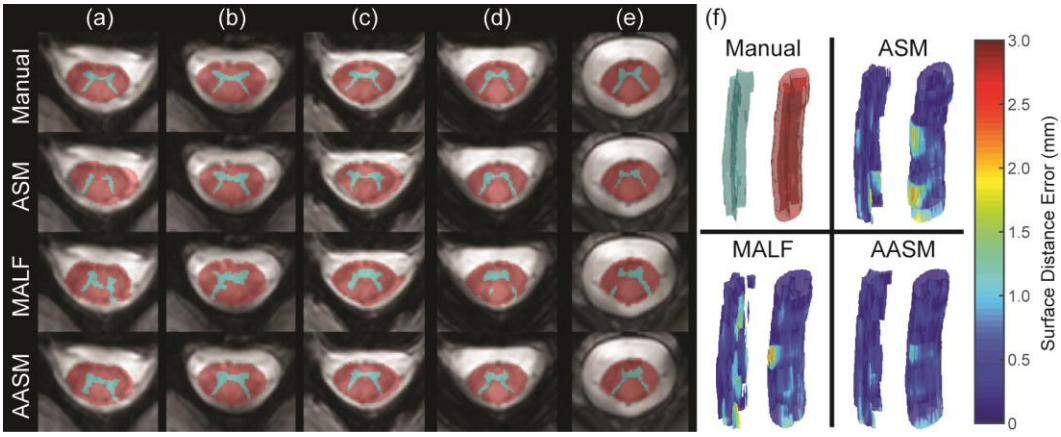

Figure VI.4. Qualitative comparison of ASM, MALF, and AASM segmentation of spinal cord. (a) - (e) demonstrate slices at five different locations (from bottom to top) on one subject. (f) illustrates the 3-D surface renderings of the segmented GM (left) and WM (right) colored in the surface distance error towards the corresponding manual segmentations.

converted the segmentation problem of WM and GM as to extract the surface of the whole SC and GM.

3.3.3. Training

Using marching squares, landmarks along the contours of manually labeled SC and GM were extracted. The correspondent landmarks of SC were acquired using the same way as the abdominal wall (212 in total). For GM, six key points along were first identified (two as the tips of the posterior horn, and four as the valley points on the left, right, anterior, and posterior side), each of the six segments in-between was then evenly resampled with 32 landmarks (192 in total). Active shape models covering 99% of the total variances, and local appearance models using intensity gradient profiles of 2 pixels (5 pixels in total) were built for both structures.

3.3.4. Testing

For each target slice, a MALF process similar to that of the abdominal wall segmentation above was performed. The probability map of GM was generated by normalizing the GM label probability, while that of SC was normalized the sum of WM and GM label probability. For both structures, 10 iterations of LS evolution were performed within masked regions (larger than zero probability) to drive the landmark updates towards the +3 level set. The time step, evolution coefficient, and smoothness factor and was set to 0.01, 10000, and 0.0001, respectively. The active shape searching range was 3 pixels along each side of the normal direction. The shape updates were constrained within ± 0.5 and ± 1 standard derivations of the Eigenvalues for SC and GM, respectively over 100 iterations at a single level.

After the SC and GM were segmented, WM was derived as the region by excluding GM from SC. The WM and GM labels were then transferred back to the original target space.

3.3.5. Results

The segmentation results were validated against the manual labels on the 28 testing volumes using DSC, MSD, and HD with comparison to results using ASM and MALF individually. MALF failed to preserve the GM shape. ASM could fall into local minimums and generated outliers that not matched

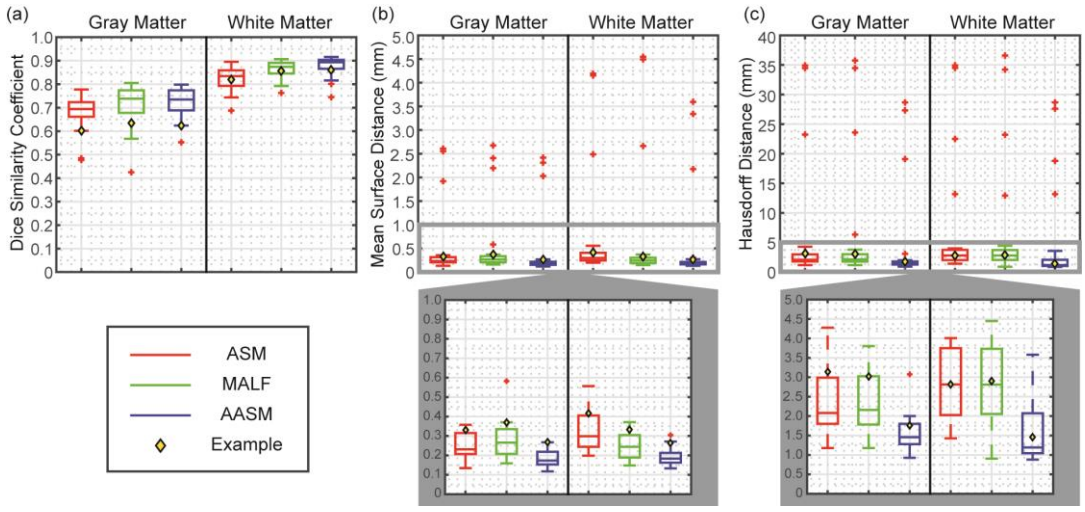

Figure VI.5. Quantitative comparison of ASM, MALF, and AASM segmentation of spinal cord in terms of DSC, MSD and HD. Note that additional zoomed-in boxplots are generated for MSD and HD to compare the results in a limited range. The yellow diamond marks indicate the subject demonstrated in Figure VI.4.

with the underlying structures. AASM captured the shape of WM and GM more robustly even with the presence of lesions (**Figure VI.4a-4e**), and yielded WM and GM surfaces smoothly along the whole volume with less surface distance error (**Figure VI.4f**). Quantitatively across 28 subjects, AASM significantly ($p < 0.005$ in cases using single-tail *t*-tests) increased the mean DSC value by 0.01, decreased the mean MSD by 0.11 mm, and decreased the mean HD by 1.43 mm comparing to the best case of ASM and MALF (**Figure VI.5**).

3.4. Parameter Sensitivity

3.4.1. Experimental Design

Here we revisited the toy example presented above to examine the parameter sensitivity of our proposed method. Using the same 100 observations, the performances of ASM and AASM were evaluated given the change on the following cases one at a time. Note that all other parameters remained as the baseline configuration in section 3.1 if not changed.

(1) The mean shape model was initialized with an offset (-10 ~ + 10 voxels along both x and y axes) to the centroid of the image.

(2) The local search range was set to range from 2 to 40 voxels along each side of the normal direction of the landmarks. Note the local appearance model was trained the intensity gradient profile length from 1 to 20 voxels correspondently along each side of the normal direction of the landmarks.

(3) The probability map was derived from the ground truth by applying a Gaussian smoothing kernel with the standard deviation ranging from 1 to 20 voxels. Note the size of the kernel on each dimension was parameterized as 4 times the standard deviation plus 1. Note that ASM configuration stayed as baseline.

3.4.2. Results

The average DSC and MSD performances for ASM and AASM across the 100 observations were collected for each specific parameter setting (**Figure VI.6**). ASM was very sensitive when the model was not initialized appropriately, while AASM demonstrated its robustness given various initializations. When the local search range was larger than 16 voxels, ASM had more chances to fall into local minimum, and

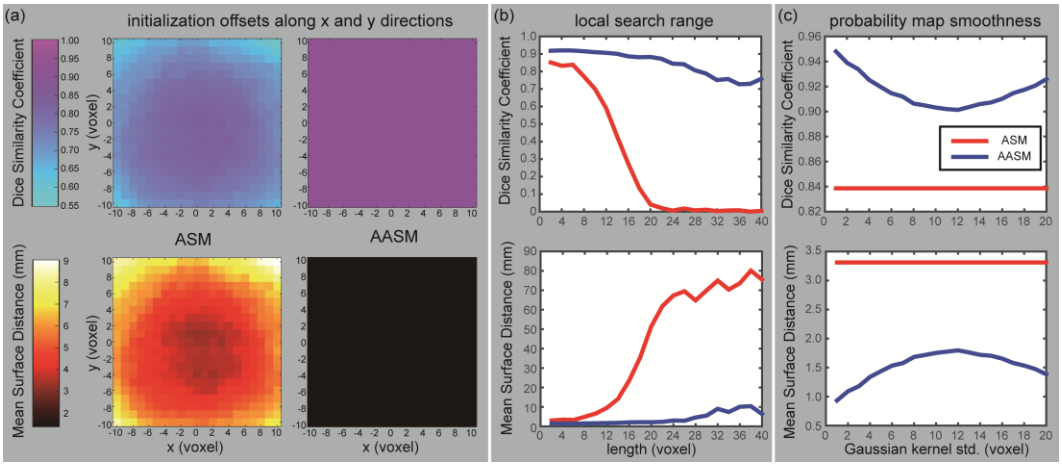

Figure VI.6. Parameter sensitivity tests to compare ASM and AASM on the toy example. DSC and MSD are collected given three sets of varying parameters: (a) the initialized mean shape position with respect to the ground truth along x and y directions, (b) the length of the local search range, and (c) the standard deviation (std.) of the Gaussian kernel applied on the ground truth, based on which the probability map used by AASM is generated (no change on ASM).

thus presented catastrophic performances, while AASM remained stable given its property of global optimization. Maximum difference of 0.05 in DSC and 0.09 mm in MSD were observed across the tested range of the smoothness of the probability map for AASM, where consistent superiority were presented over ASM.

4. Discussion and Conclusion

In this study, we proposed an automatic framework (AASM) that coherently integrates three modern image segmentation techniques including ASM, MALF, and LS. Great synergies were found within this framework, where (1) ASM and MALF used the same training datasets to generate statistical models, and probabilistic atlases, respectively, (2) LS using the CV speed function can be directly applied to the probability map generated from MALF, and (3) the region-based LS evolution extends the range of correspondent landmark search of ASM. Using AASM, challenging segmentation problems can be benefitted from the shape regularization and topology preservation of ASM, contextual robustness of MALF, and global optimization of region-based LS. On 20 abdominal CT scans, we presented the first automatic segmentation approach to extract the outer and inner surfaces of the whole abdominal wall covering thoracic, abdominal, and pelvic regions, and thus enabled subcutaneous and visceral fat measurement with high correlation to the measurement derived from manual segmentation. On 28 3T MR scans of cervical SC, we demonstrated robust WM and GM segmentation with the presence of MS lesions. It is worth to note that with minor post-processing (3-D smoothing), our SC segmentation approach improved the start-of-the-art method [192] by 25% in DSC on the same datasets. We see huge opportunities to adapt the proposed method to other anatomical structures in medical images, whose complexity cannot be easily handled by ASM, MALF, or LS individually.

AASM has many parameters to configure as it combines three image segmentation techniques. While universally ideal configuration can be hardly found, robustness can be achieved for individual applications with empirical parameter settings, especially the balance between the length of local search range and the number of LS evolution. For example, the LS evolution should dominate the shape updates

if the local contexts around the structure are ambiguous (see the spinal cord application), while the active shape search is preferred if the local appearances are uniquely identifiable (see the abdominal wall application). In any cases, the structural shapes are properly regularized within the proposed framework.

On both clinical datasets, the larger variations were observed over the secondary structures than the structure of interest, i.e., the abdominal wall and spinal cord, along the cranial-caudal direction. In addition, slices at different locations presented various shapes, appearances, and contexts. Therefore, we performed slice-wise AASM segmentation by using target-specific statistical shape and appearance models to capture the desirable variations. The target-specific selection of the training datasets has a substantial impact on ASM and MALF, and thus the overall performance of AASM. Investigation in the classification system for clustering similar datasets can be beneficial for further improvement. On the other hand, the 2-D AASM implementation substantially simplifies the assignment of correspondent landmarks across the training datasets. The 3-D extension of AASM framework is straightforward, while the landmark correspondences need to be properly handled [111, 193]. Another perspective of future work can focus on multi-region.

PART 3

ABDOMINAL ORGANS

In this part, we concentrate on the automatic segmentation abdominal organs including (1) spleen, (2) right kidney, (3) left kidney, (4) gallbladder, (5) esophagus, (6) liver, (7) stomach, (8) aorta, (9) inferior vena cava, (10) portal and splenic vein, (11) pancreas, (12) right adrenal gland, and (13) left adrenal gland. We mainly use a non-parametric segmentation approach, i.e., multi-atlas label fusion, because of its relatively stronger robustness compared to other methods. In abdomen, we consider the image registrations errors as the main challenge for multi-atlas spleen segmentation caused by the huge variations among the clinically acquired abdominal CT scans, and propose to relieve the problem from the following perspectives. Specifically, we first focus on spleen, and use the shape constraint to regularize the spleen segmentation with a level-set-based active shape model, where regions don't fit well in an adapted spleen shape are suppressed (**Chapter VII**). Then, we propose that appropriate atlas selection before the label fusion process can significantly reduce the impact of non-robust registrations, where the learned knowledge on target context and the intermediate fusion results are used to select the best possible atlases for fusion. Integrated with the context learned atlas selection, a generic multi-atlas framework is established for robust segmentation of multiple abdominal organs (**Chapter VIII**). Next, we evaluate the clinical efficacy of our automatic segmentation. A pipeline combining computer-assisted segmentation and manual outlier correction is proved more robust and efficient for spleen volume estimation than traditional clinical approaches on a longitudinal dataset (**Chapter IX)**. Registrations are extensively investigated given its importance for atlas-based segmentation. Through over 100,000 hours of CPU time for validating six methods on abdominal organ registrations across 100 datasets, we provide recommendation for registration users and suggest directions for registration developers (**Chapter X**). Random forest technique for organ localization is investigated and integrated with the multi-atlas

segmentation framework. Organ-wise multi-atlas labeling performed on the localized region of interest provides consistent improvement in segmentation performances (**Chapter XI**).

CHAPTER VII

SHAPE-CONSTRAINED MULTI-ATLAS SEGMENTATION OF SPLEEN IN CT

1. Introduction

Automated spleen segmentation on clinical CT data is a challenging problem due to the complicity and the variability of abdominal anatomy across populations. Multi-atlas segmentation is a potential approach that can provide robust segmentation of the spleen. Voting (e.g., majority vote (MV [13]), locally weighted vote (LWV [133])) and statistical fusion (e.g., STAPLE [125]) algorithms have been developed to improve the efficacy of combing potentially conflicting labels from multiple atlases. However, registrations of abdominal CT images are problematic given large-scale difference in abdominal anatomy and similar intensities between organs. The atlas labels with problematic registrations may misalign with the target structure, and/or include various redundant structures, and thus substantially undermine the fusion results even when additional intensity information is considered (**Figure VII.1**).

We posit that the integration of shape information into the label fusion framework can provide more accurate fusion in cases of problematic registrations between atlas and target images. Although extensively used, point-based shape models have practical limitations as they require manual annotation of corresponding landmarks. Alternatively, an implicit shape model [113-115] has been used to characterize shape variations by voxel-wise high dimensional signed distance function (SDF), where each volume with its all voxels (instead of several landmarks) is considered as a single observation for the shape training. The implicit shape model has been successfully applied to some level set approaches of anatomical structures to constrain the shape of the segmentation [115]. Herein, we propose a multi-atlas segmentation framework that integrates the implicit shape model to regularize the fusion results.

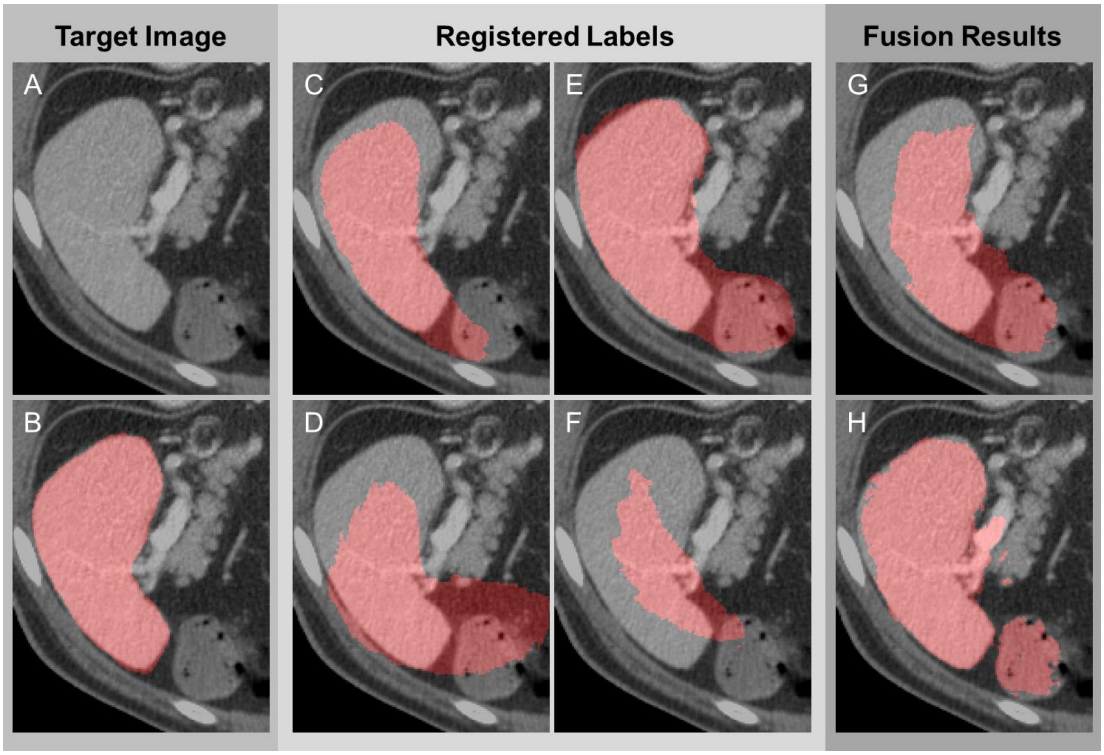

Figure VII.1. Challenges of atlas-based spleen segmentation. (A) The intensity image of the target. (B) The manually labeled ground truth of the target. (C) – (F) The atlas labels registered to the target space. (G) The fusion estimate by majority vote. (H) The fusion estimate by locally weighted vote.

2. Methods and Results

2.1. Spleen Localization

In our previous spleen approach, we used a regression method, i.e., regression forests region recognition (RFRR [194]), to localize the spleen region with a bounding box and demonstrated that multi-atlas segmentation of spleen provides better estimate on the reduced region of interest. In this study, all intensity and label images referred below were cropped in terms of the identified bounding boxes. Here we omit the mechanism of RFRR for brevity.

2.2. Pose-Free Implicit Shape Model

All atlas label images are co-registered into the same space with 7 degrees of freedom (DOF). The binary labels are transformed into high dimensional SDF to implicitly represent the shape of spleen, and then collected in an $M \times N$ observation matrix A, where M is the number of observations, and N is the number of voxels. A Principle Component Analysis (PCA) procedure is used to extract the variability of spleen shape, where the mean shape $\bar{\Phi}$ is the voxel-wise average across observations,

$$\bar{\Phi} = \frac{\sum_m A_{m\cdot}}{M} \tag{7.1}$$

and the modes of variation are provided by singular value decomposition (SVD) on the covariance of the

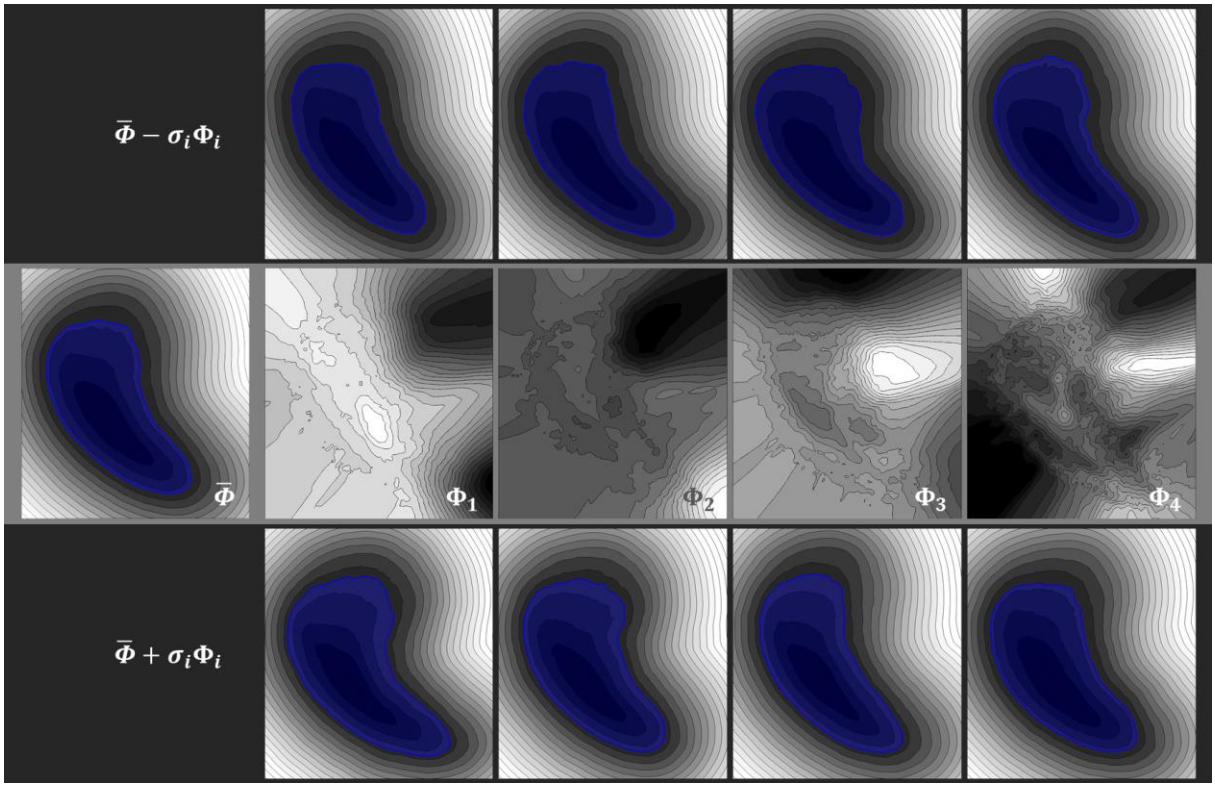

Figure VII.2. Pose-free implicit parametric shape model. The shape model is represented by signed distance function (SDF) of each voxel over the whole volume. The region within the zero level set (highlighted in blue) is considered as the binary shape representation. The second row illustrates the mean and the first four modes of variation of the shape model. The first and the third row present the specific shapes parameterized by the square root of the eigenvalues.

centered observation matrix,

$$U\Sigma U^T = \frac{1}{M}(A - \bar{\Phi})^T(A - \bar{\Phi}) \tag{7.2}$$

where Σ is a diagonal matrix with its diagonal values as the eigenvalues λ_i, and U denotes an eigenspace with each column as an eigenvector, i.e., one mode of variation Φ_i, associated with the corresponding eigenvalue λ_i (**Figure VII.2**). The value of the eigenvalue indicates the dominance of its associated mode of variation, while the modes with relatively small eigenvalues are usually ignored due to their limited variances provided.

Given the implicit shape model, a specific shape can be then characterized by the combination of the modes of variations on the basis of the mean shape,

$$\Phi \equiv \Phi(\boldsymbol{\omega}) = \bar{\Phi} + \sum_i \omega_i \Phi_i \tag{7.3}$$

where ω_i denotes the shape parameter associated with its mode of variation.

2.3. Shape-Constrained Multi-Atlas Segmentation Framework

2.3.1. Initial Estimation

We initialize with a regular fusion of the registered atlas labels via locally weighted vote (LWV). In particular, we define the weight on voxel i between the jth registered atlas image S_j and the target image T in terms of intensity similarity in a $3 \times 3 \times 3$ neighborhood Ω,

$$w_{ij} = \exp\left(-\frac{\sum_{k\epsilon\Omega}\|S_{kj} - T_k\|^2}{\sigma}\right) \tag{7.4}$$

where σ is a parameter that controls the de-weighting degree in terms of the local dissimilarity. Comparing to MV, LWV tends to capture a more complete spleen volume even though some regions are not covered by the majority of atlas labels.

2.3.2. Shape Registration

The pose-free implicit shape model is then transformed into the target space based on the registration between the binary image of the mean shape and that of the current segmentation. We found

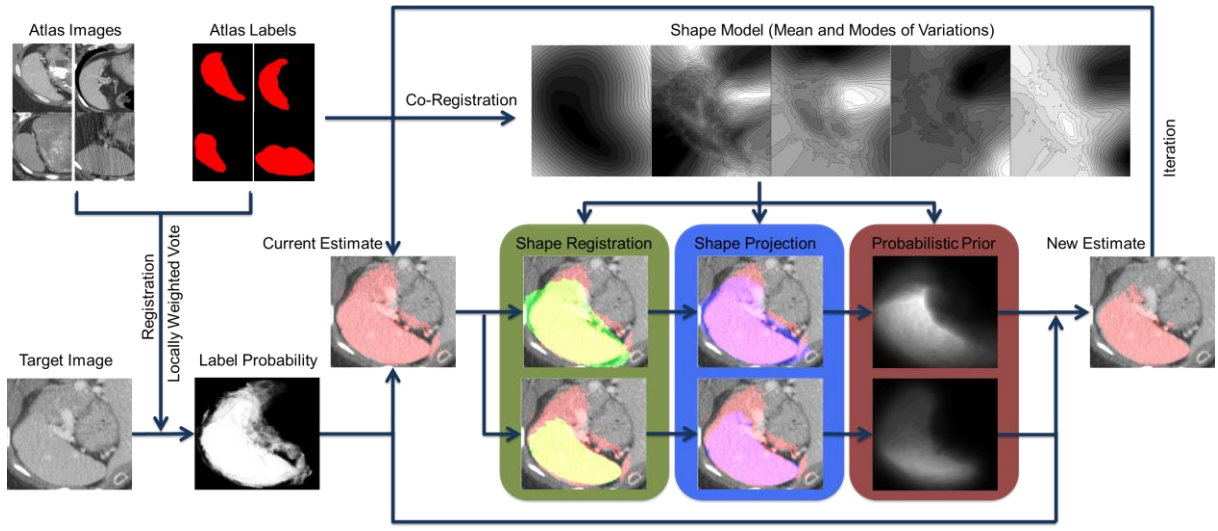

Figure VII.3. Flowchart of the proposed method. The atlas labels are co-registered to construct a pose-free implicit parametric shape model, including the mean and the modes of variation of the spleen shape. The atlas images are registered to the target image, based on which the atlas labels are propagated to the target space. The locally weighted vote yields the initial fusion result of the registered atlas labels weighted by the local intensity similarity between the registered atlas images and the target image in the form of a fuzzy estimate of label probability and a binary estimate of spleen segmentation. The binary image of the mean shape from the pre-constructed spleen shape model is registered to that of the current estimate with two distinct effective ranges, i.e., (1) the whole volume of both image and (2) the mean shape region, of the similarity metric of registration so that the pose-free shape model is transformed into the target space. The current estimate of the spleen is then projected to the two registered shape models. The shape projections are converted into probabilistic priors to adjust the label probability from locally weighted vote, and then generate a new estimate of the spleen. The estimate can be refined with iterative adjustment.

that a single registration on binary images is practically error-prone due the existence of massive missing/redundant structures. Therefore, we apply two registrations between these binary images with two distinct effective ranges, i.e., (1) the whole volume of both image and (2) the mean shape region, of the similarity metric of registration, so that the two sets of registered mean shapes tend to capture the

outer and inner boundary of the current estimate, respectively. The registrations use normalized correlation criterion as the similarity metric with 7 DOF.

2.3.3. Shape Projection

The segmentations can then be projected to the registered shape model based on the mean shape registration which effectively constrains the estimate within the shape model. In particular, based on each set of two registrations, the pose-free shape model is transformed into the target space. The current estimate of the spleen is converted into SDF, i.e., Φ_c, and projected to two registered shape models.

$$\hat{\boldsymbol{\omega}} = U(\Phi_c - \bar{\Phi}) \tag{7.5}$$

where $\hat{\boldsymbol{\omega}}$ is the projected shape parameter, which is then used to reconstruct the projected shape,

$$\hat{\Phi} = U^T\hat{\boldsymbol{\omega}} + \bar{\Phi} \tag{7.6}$$

In implementation, we restrict the absolute value of projected shape parameter to be no larger than the square root of its corresponding eigenvalue to preserve a reasonable reconstructed spleen shape.

2.3.4. Shape-Probability Conversion

The output of the shape projection can be considered as the representation of how far voxels are inside/outside the projected spleen shape, which can be converted into a shape prior of how likely the voxels represent the spleen via an exponential function,

$$\mathrm{P} = \exp(-\alpha\hat{\Phi}) \tag{7.7}$$

where α indicates the steepness of the conversion from SDF to probability.

2.3.5. Iterative Refinement

The shape probabilistic priors, along with the label probability provided by LWV, are used to generate a new estimate of the spleen, and the fusion estimate can be refined with iterative adjustment.

Please refer to **Figure VII.3** to the detailed flowchart of the proposed framework.

Table VII.1. Error metrics via different fusion methods.

Metrics	MV	LWV	Proposed Method
DSC	0.69±0.17	0.77±0.13	0.83±0.08
MSD (mm)	8.01±6.12	7.49±7.66	3.48±1.88
HD (mm)	33.81±19.00	45.88±27.84	22.67±8.23

2.4. Data and Validation

Under an Institutional review board waiver, 25 portal venous phase contrast-enhanced CT abdomen scans were randomly selected from a larger ongoing colorectal cancer chemotherapy trial. Images were approximately 512 x 512 x 152 with a resolution of 0.7 x 0.7 x 3.0 mm. Scans with poor contrast bolus timing (i.e. not portal venous phase) or aberrant patient positioning were excluded, leaving 20 scans for analysis. Spleens were manually labeled by an experienced radiologist on a volumetric basis using the MIPAV software (NIH, Bethesda, MD [153]). All intensity and label images were cropped in terms of the bounding boxes around the localized spleen region as described in section 2.1.

All subjects are processed in a leave-one-out cross validation framework. Comparing to a baseline LWV method, our proposed shape-constrained method achieves significantly higher Dice similarity coefficient (DSC [146]) ($\Delta = 0.06, p < 0.01$), significantly lower symmetric mean surface distance (MSD) ($\Delta = -4.01\ mm, p < 0.01$), and significantly lower symmetric Hausdorff surface distance (HD) ($\Delta = -23.21\ mm, p < 0.01$) (**Table VII.1**) verified by Wilcoxon signed rank tests. Our shape-constrained method presents a substantial effect in correcting the spleen shape (**Figure VII.4**).

3. Discussion

Spleen segmentation in CT images via multi-atlas segmentation can be detrimentally impacted by problematic registrations between atlases and target images. While the pose-free implicit shape model is not difficult to construct, challenges lie in the model-estimate alignment and shape-probability conversion. We used two sets of registrations with distinct effective ranges of similarity metric between the fusion estimate and the mean shape of shape model to provide the inner and outer boundary of the spleen shape to compensate for the uncertainty of the correct pose alignment. We project the fusion

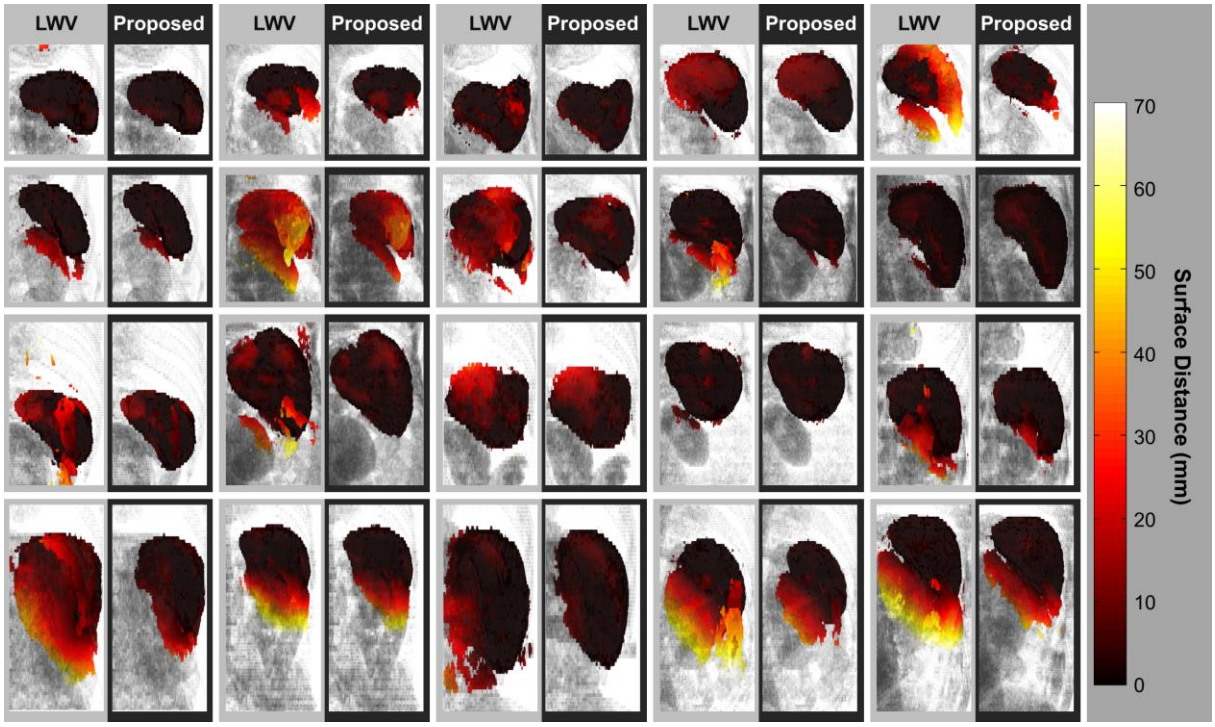

Figure VII.4. Fusion results by locally weighted vote and the proposed shape-constrained method on 20 subjects. The results of the two methods are placed side-by-side for each subject for comparison. The background rendering provides a reference of the surrounding anatomy (ribs, kidney, etc.). The rendering of the spleen segmentation is colored in terms of the surface distance from the estimate to the ground truth. Note that we use symmetric (the average of bi-directional) surface distance as the error metrics for validation in Table 1, but the one-way surface distance here for ease of visualization.

estimate to the registered shape model, which yields finer matchup with the fusion estimate under shape constraint. The projected shape represented in SDF is converted into the shape probabilistic prior through an exponential function, and thus regularizes the fusion estimate by shape information.

We note the pose-free implicit shape model can be built upon the atlas labels for label propagation with no additional data required. This presents opportunities to combine the proposed framework with the state-of-art label fusion methods (e.g., [195]) to yield shape-constrained fusion estimate. In addition, the usage of the shape information, beyond a simple probabilistic prior, requires further study.

CHAPTER VIII

EFFICIENT MULTI-ATLAS ABDOMINAL SEGMENTATION ON CLINICALLY ACQUIRED CT WITH SIMPLE CONTEXT LEARNING

1. Introduction

The human abdomen is an essential, yet complex body space. Computed tomography (CT) scans are routinely obtained for the diagnosis and prognosis of abdomen-related disease. Automated segmentation of abdominal anatomy may improve patient care by decreasing or eliminating the subjectivity inherent in traditional qualitative assessment. In large-scale clinical studies, efficient segmentation of multiple abdominal organs can also be used for biomarker screening, surgical navigation, and data mining.

Atlas-based segmentation provides a general-purpose approach to segment target images by transferring information from canonical atlases via registration. When adapting to abdomen, the variable abdominal anatomy between individuals (e.g., weight, stature, age, disease status) and within individuals (e.g., pose, respiratory cycle, clothing) can lead to substantial registration errors (**Figures VIII.1, VIII.2**). Previous abdominal segmentation approaches have used single probabilistic atlases constructed by co-registering atlases to characterize the spatial variations of abdominal organs [16, 18]; statistical shape models [23, 196] and / or graph theories [24, 27] have been integrated to refine the segmentation using probabilistic atlases. Multi-atlas segmentation (MAS), on the other hand, is a technique that has been proven effective and robust in neuroimaging by registering multiple atlases to the target image separately, and combining voxel-wise observations among the registered labels through label fusion [133]. Recently, Wolz et al. applied MAS to the abdomen using locally weighted subject-specific atlas [28]; yet the segmentation accuracies were inconsistent. We posit that the efficiency of atlas selection for abdominal

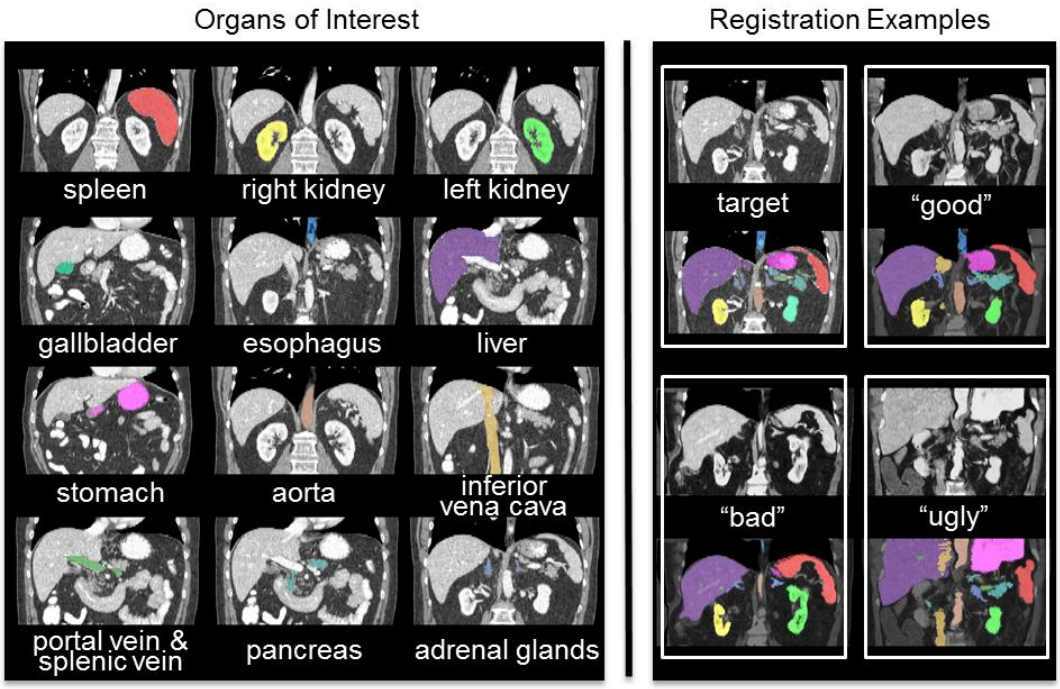

Figure VIII.1. Twelve organs of interest (left) and registration examples of variable qualities for one target image (right). Note that the "good", "bad", and "ugly" registration examples were selected regarding the organ-wise correspondence after the atlas labels were propagated to the target image.

MAS requires further exploration in the context of substantial registration errors, especially on clinically acquired CT.

The selective and iterative method for performance level estimation [127] (SIMPLE) algorithm raised effective atlas selection criteria based on the Dice similarity coefficient [146] overlap with intermediate voting-based fusion result, and addressed extensive variation in prostate anatomy to reduce the impact of outlier atlases. In [42], we generalized a SIMPLE theoretical framework to account for exogenous information through Bayesian priors – referred to as context learning; the newly presented model selected atlases more effectively for segmenting spleens in metastatic liver cancer patients. A further integration with joint label fusion (JLF [134]) addressed the label determination by reducing the correlated errors among the selected atlases, and yielded a median DSC of 0.93 for spleen segmentation.

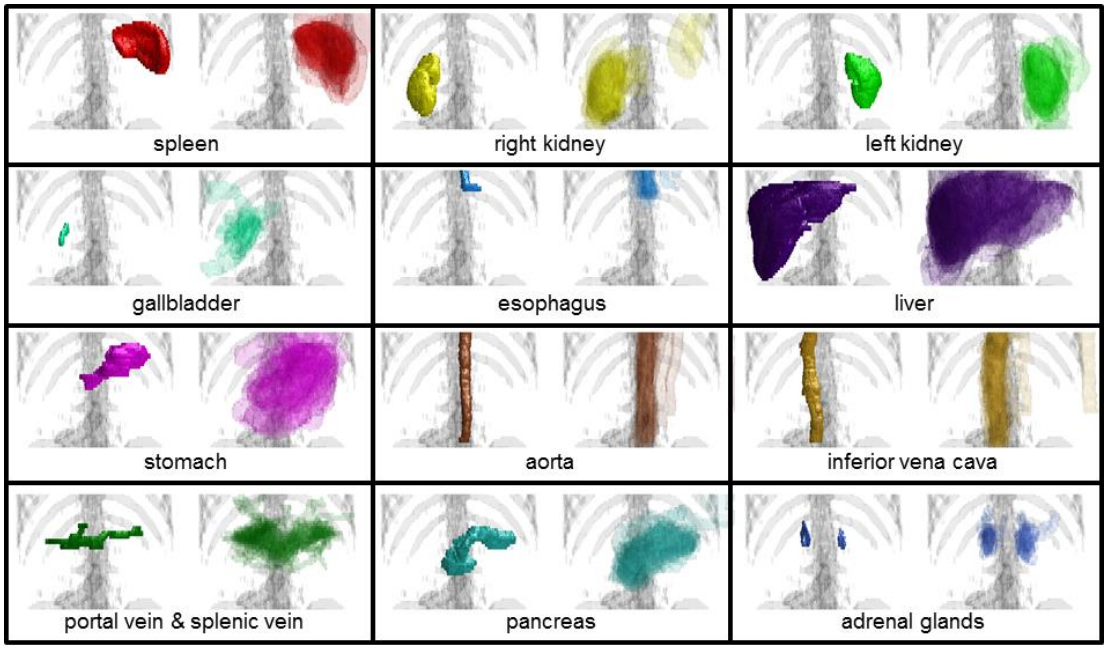

Figure VIII.2. Organ-wise examples of variations after non-rigid registrations. For each panel, the target manual segmentation is on the left, the 30 registered labels are semi-transparently overlaid on the right.

Herein, we propose an efficient approach for segmenting 12 abdominal organs of interest (**Figure VIII.1**) in 75 metastatic liver cancer patients and 25 ventral hernia patients on clinically acquired CT. Based on the re-derived SIMPLE framework [42], we construct context priors, select atlases, and fuse estimated segmentation using JLF for each organ individually, and combine the fusion estimates of all organs into a regularized multi-organ segmentation using graph cut [101] (**Figure VIII.3**). The segmentation performances are validated with other MAS approaches, including majority vote (MV), SIMPLE, JLF, and the Wolz approach. This work is an extension of previous theoretical [42] and empirical [43] conference papers and presents new analyses of algorithm performance and parameter sensitivity.

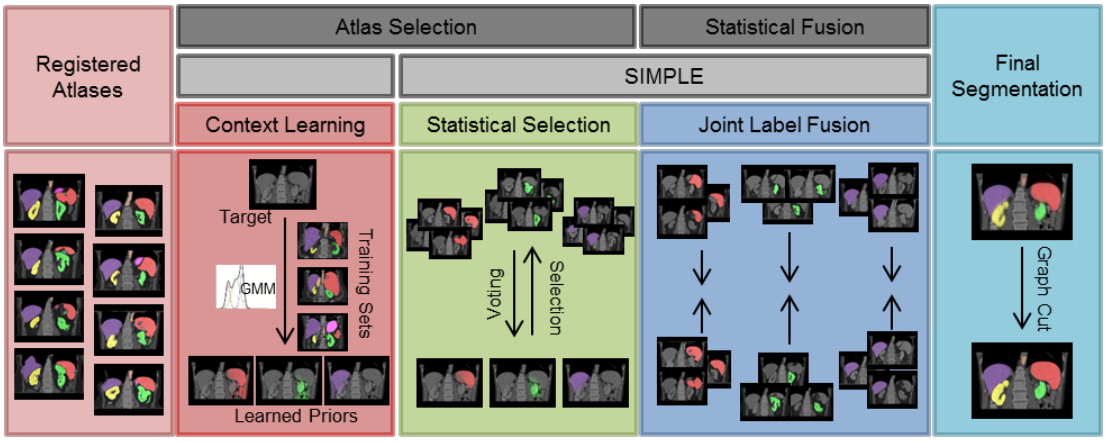

Figure VIII.3. Flowchart of the proposed method. Given registered atlases with variable qualities, atlas selection and statistical fusion are considered as two necessary steps to obtain a reasonable fusion estimate of the target segmentation. The SIMPLE algorithm implicitly combines these two steps to fusion selected atlases; however, more information can be incorporated to improve the atlas segmentation, and a more advanced fusion technique can be used after the atlases are selected. We propose to (1) extract a probabilistic prior of the target segmentation by context learning to regularize the atlas selection in SIMPLE for each organ, (2) use Joint Label Fusion to obtain the probabilistic fusion estimate while characterizing the correlated errors among the selected organ-specific atlases, and render the final segmentation for all organs via graph cut.

2. Theory

We re-formulate the SIMPLE algorithm from the perspective of Expectation-Maximization (EM) while focusing on the atlas selection step. In this principled likelihood model, the Bayesian prior learning from context features (e.g., intensity, gradient) is considered as exogenous information to regularize the atlas selection.

2.1. Statistical SIMPLE Model

Consider a collection of R registered atlases with label decisions, $\boldsymbol{D} \in \boldsymbol{L}^{N\times R}$, where N is the number of voxels in each registered atlas, and $\boldsymbol{L} = \{0,1,\dots,L-1\}$ represents the label sets. Let $\boldsymbol{c} \in \boldsymbol{S}^{R}$, where $\boldsymbol{S} = \{0,1\}$ indicates the atlas selection decision, i.e., 0 – ignored, and 1 – selected. Let i be the

index of voxels, and j of registered atlases. We propose a non-linear rater model, $\theta \in \mathbb{R}^{R\times 2\times L\times L}$, that considers the two atlas selection decisions. Let the ignored atlases be no better than random chance, and the selected atlases be slightly inaccurate with error factors $\boldsymbol{\epsilon} \in \mathbf{E}^{\mathrm{R}\times 1}$, where $\mathbf{E} \in (0, \frac{L-1}{L})$. Thus

$$\theta_{j0s's} = \frac{1}{L}, \quad \forall s' \; ; \; \theta_{j1s's} = \begin{cases} 1-\epsilon_j, & s' = s \\ \dfrac{\epsilon_j}{L-1}, & s' \neq s \end{cases} \tag{8.1}$$

where each element $\theta_{jns's}$ represents the probability that the registered atlas j observes label s' given the true label is s and the atlas selection decision is n with an error factor ϵ_j if selected, – i.e., $\theta_{jns's} \equiv f\left(D_{ij} = s' \middle| T_i = s, c_j = n, \epsilon_j\right)$.

Following [125], let $\boldsymbol{W} \in \mathbb{R}^{L\times N}$, where $W_{si}^{(k)}$ represents the probability that the true label associated with voxel i is label s at the k$^{\mathrm{th}}$ iteration. Using Bayesian expansion and conditional inter-atlas independence, the E-step can be derived as

$$W_{si}^{(k)} = \frac{f(T_i = s)\prod_j f\left(D_{ij}\middle| T_i = s, c_j^{(k)} = n, \epsilon_j^{(k)}\right)}{\sum_{s'} f(T_i = s')\prod_j f\left(D_{ij}\middle| T_i = s', c_j^{(k)} = n, \epsilon_j^{(k)}\right)} \tag{8.2}$$

where $f(T_i = s)$ is a voxel-wise *a priori* distribution of the underlying segmentation. Note that the selected atlases contribute to $\boldsymbol{W}$ in a similar way as globally weighted vote given the symmetric form of $\theta_{j1s's}$ as in the original SIMPLE.

In the M-step, the estimation of the parameters is obtained by maximizing the expected value of the conditional log likelihood function found in Eq. 2. For the error factor,

$$\begin{aligned} \epsilon_j^{(k+1)} &= \arg\max_{\epsilon_j} \sum\nolimits_i E\left[\ln f\left(D_{ij}\middle| T_i, c_j^{(k)}, \epsilon_j\right) | \boldsymbol{D}, {c_j}^{(k)}, \epsilon_j^{(k)}\right] \\ &= \arg\max_{\epsilon_j} \sum\nolimits_{s'} \sum\nolimits_{i:D_{ij}=s'} \sum\nolimits_s W_{si}^{(k)} \ln \theta_{jc_j^{(k)}s's} \equiv L_{\epsilon_j} \end{aligned} \tag{8.3}$$

Consider the binary segmentation for simplicity, let $M_{TP} = \sum_{i:D_{ij}=1} W_{1i}^{(k)}$, $M_{FP} = \sum_{i:D_{ij}=1} W_{0i}^{(k)}$, $M_{FN} = \sum_{i:D_{ij}=0} W_{1i}^{(k)}$, $M_{TN} = \sum_{i:D_{ij}=0} W_{0i}^{(k)}$, and $\mathrm{M}_T = M_{TP} + M_{TN}$, $M_F = M_{FP} + M_{FN}$. After taking partial derivative of L_{ϵ_j},

$$\epsilon_j^{(k+1)} = \frac{M_F}{M_T + M_F}, i.e., 1 - \epsilon_j^{(k+1)} = \frac{M_T}{M_T + M_F} \tag{8.4}$$

Then for the atlas selection decision

$$\begin{aligned} c_j^{(k+1)} &= \arg\max_{c_j} \sum_i E\left[\ln f\left(D_{ij}\middle|T_i, c_j, \epsilon_j^{(k+1)}\right)|\boldsymbol{D}, {c_j}^{(k)}, \epsilon_j^{(k+1)}\right] \\ &= \arg\max_{c_j} \sum_{s'} \sum_{i:D_{ij}=s'} \sum_s W_{si}^{(k)} \ln \theta_{jc_j s's} . \end{aligned} \tag{8.5}$$

Given the intermediate truth estimate $W_{si}^{(k)}$, $c_j^{(k+1)}$ can be maximized by evaluating each 0/1 atlas selection separately. Note the selecting/ignoring behavior in Eq. 5 is parameterized with the error factor ϵ_j, and thus affected by the four summed values of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) as in Eq. 4. Typical practice for a fusion approach might use the prior probability, $f(T_i = s)$, to weight by expected volume of structure. With outlier atlases, one could reasonably expect a much larger region of confusion (i.e., non "consensus"[138]) than true anatomical volume. Hence, an informed prior would greatly deemphasize the TN and yield a metric similar to DSC. Therefore, we argue that SIMPLE is legitimately viewed as a statistical fusion algorithm that is approximately optimal for the non-linear rater model proposed in Eq. 1.

2.2. Context Learning

Different classes of tissues in CT images can be characterized with multi-dimensional Gaussian mixture models using intensity and spatial "context" features. On a voxel-wise basis, let $\boldsymbol{v} \in \mathbb{R}^{d\times 1}$ represent a d dimensional feature vector, $m \in \boldsymbol{M}$ indicate the tissue membership, where $\boldsymbol{M} = \{1, \dots, M\}$ is the set of possible tissues, and typically, a superset of the label types, i.e., $\boldsymbol{M} \supseteq \boldsymbol{L}$. The probability of the observed features given the tissue type is t can be represented with the mixture of N_G Gaussian distributions,

$$f(\boldsymbol{v}|m = t) = \sum_{k=1}^{N_G} \frac{\alpha_{kt}}{(2\pi)^{\frac{d}{2}}|\boldsymbol{C}_{kt}|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(\boldsymbol{v} - \boldsymbol{\mu}_{kt})^T \boldsymbol{C}_{kt}^{-1}(\boldsymbol{v} - \boldsymbol{\mu}_{kt})\right] \tag{8.6}$$

where $\alpha_{kt} \in \mathbb{R}^{1\times1}$, $\boldsymbol{\mu}_{kt} \in \mathbb{R}^{d\times1}$, and $\boldsymbol{C}_{kt} \in \mathbb{R}^{d\times d}$ are the unknown mixture probability, mean, and covariance matrix to estimate for each Gaussian mixture component k of each tissue type t by the EM algorithm following [107]. This context model can be trained from datasets with known tissue separations.

The tissue likelihoods on an unknown dataset can be inferred by Bayesian expansion and can use a flat tissue membership probability from extracted feature vectors.

$$f(m = t|\boldsymbol{v}) = \frac{f(\boldsymbol{v}|m = t)f(m = t)}{\sum_{t'} f(\boldsymbol{v}|m = t')f(m = t')} = \frac{f(\boldsymbol{v}|m = t)}{\sum_{t'} f(\boldsymbol{v}|m = t')} \tag{8.7}$$

Consider a desired label s as one tissue type t, and thus $f(T_i = s) \equiv f(m = t|\boldsymbol{v})$, the Bayesian prior learning from context features serves to regularize the intermediate fusion estimate in Eq. 3, and hence the atlas selection.

3. Methods and Results

3.1. Data

Under Institutional Review Board (IRB) supervision, the first-session of abdomen CT scans of 75 metastatic liver cancer patients were randomly selected from an ongoing colorectal cancer chemotherapy trial, and an additional 25 retrospective scans were acquired clinically from post-operative patients with suspected ventral hernias. The 100 scans were captured during portal venous contrast phase with variable volume sizes (512 x 512 x 33 ~ 512 x 512 x 158) and field of views (approx. 300 x 300 x 250 mm^3 ~ 500 x 500 x 700 mm^3). The in-plane resolution varies from 0.54 x 0.54 mm^2 to 0.98 x 0.98 mm^2, while the slice thickness ranges from 1.5 mm to 7.0 mm. Twelve abdominal organs were manually labeled by two experienced undergraduate students, and verified by a radiologist on a volumetric basis using the MIPAV software (NIH, Bethesda, MD [153]). All images and labels were cropped along the cranio-caudal axis with a tight border without excluding liver, spleen, and kidneys before any processing (following [28]).

3.2. General Implementation

We used 10 subjects to train context models for 15 tissue types, including twelve manually traced organs, and three automatically retrieved tissues (i.e., muscle, fat, and other) using intensity clustering and excluding the traced organ regions. Six context features were extracted, including intensity, gradient, and local variance, and three spatial coordinates with respect to a single landmark, which was loosely identified as the mid-frontal point of the lung at the plane with the largest cross-sectional lung area (see rendering in **Figure VIII.8**). We specified the number of components of Gaussian mixture model, $N_G = 3$. For each organ, the foreground and background likelihoods were learned from the context models based on the context features on target images, and used as a two-fold spatial prior to regularize the organ-wise SIMPLE atlas selection. We constrained the number of selected atlases as no less than five and no larger than ten.

When using JLF on the selected atlases for each organ, we specified the local search radii (in voxel) as $3 \times 3 \times 3$, the local patch radii (in voxel) as $2 \times 2 \times 2$, and set the intensity difference mapping parameter, and the regularization term as 2 and 0.1, respectively (i.e., default parameters).

Following [28, 143], we regularized the final segmentation with graph cut (GC). The GC problem is solved by maximizing the following MRF-based energy function

$$E(p) = \lambda \sum_{i \in \mathbb{I}} D_i(p_i) + \sum_{\{i,i'\} \in \mathcal{N}} V_{i,i'}(p_i, p_{i'}) \tag{8.8}$$

where i and i' are voxel indices, p represents the labeling of the final segmentation for image $\mathbb{I}$. The data term $D_i(p_i)$ characterize the probability of voxel i assigned to the label p_i; we define it as a combination of the probabilistic fusion estimate from JLF with the intensity likelihoods using 1-D context learning. The smoothness term $V_{i,i'}(p_i, p_{i'})$ penalizes the discontinuities between the voxel pair $\{i, i'\}$ in the specified neighborhood system $\mathcal{N}$; we define it as a combination of the intensity appearance with local boundary information. λ is a coefficient that weights the data term over the smoothness term; we set it as 3.3. Note that we only applied GC smoothing to large organs (i.e., spleen, kidneys, liver, stomach), and kept the JLF results for the remaining organ structures.

For the direct JLF approach, the same parameters were used as above, except that it was conducted for all organs simultaneously. For the Wolz approach, we kept 30 atlases for the global atlas selection, adjusted the exponential decay for the organ level weighting to support 10 atlases, followed [28] for voxel-wise weighting by non-local means, and used the same GC scheme as applied to our proposed method.

Note that we used the JLF [134] method in the Advanced Normalization Tools (ANTs) [197], all other algorithms, i.e., MV [124], SIMPLE [127], the Wolz approach [28], and GC [101, 143], were implemented based on the corresponding literature, and run on a 64-bit 12-core Ubuntu Linux workstation with 48G RAM.

3.3. Motivating Simulation

3.3.1. Experimental Setup

A simulation on 2-D CT slices was constructed to demonstrate and motivate the benefits of SIMPLE context learning for atlas selection and label fusion (see **Figure VIII.4**). Forty CT scans were randomly selected from the 90 subjects not used for context learning. A representative slice with the presences of all three organs, i.e., spleen, left kidney, and liver, was extracted from each scan, and considered as a target image. A hundred simulated observations were estimated by applying a random transformation model to each target slice, and considered as the atlases with different degrees of registration errors for segmenting the target.

The simulation model involved an affine followed by a non-rigid transformation. The affine transformation consisted of a rotational component as well as two translational and two scaling components, with the effect of each component drawn from a zero-mean Gaussian distribution with standard deviations of 2 degrees for the rotational component, 5 mm for the translational components and 0.2 mm for the scaling components. The non-rigid transformation used a deformation field created by sextic Chebyshev polynomials. The Chebyshev coefficients for each grid location were randomly generated from a standard normal distribution, on the top of which, two additional factors to control the

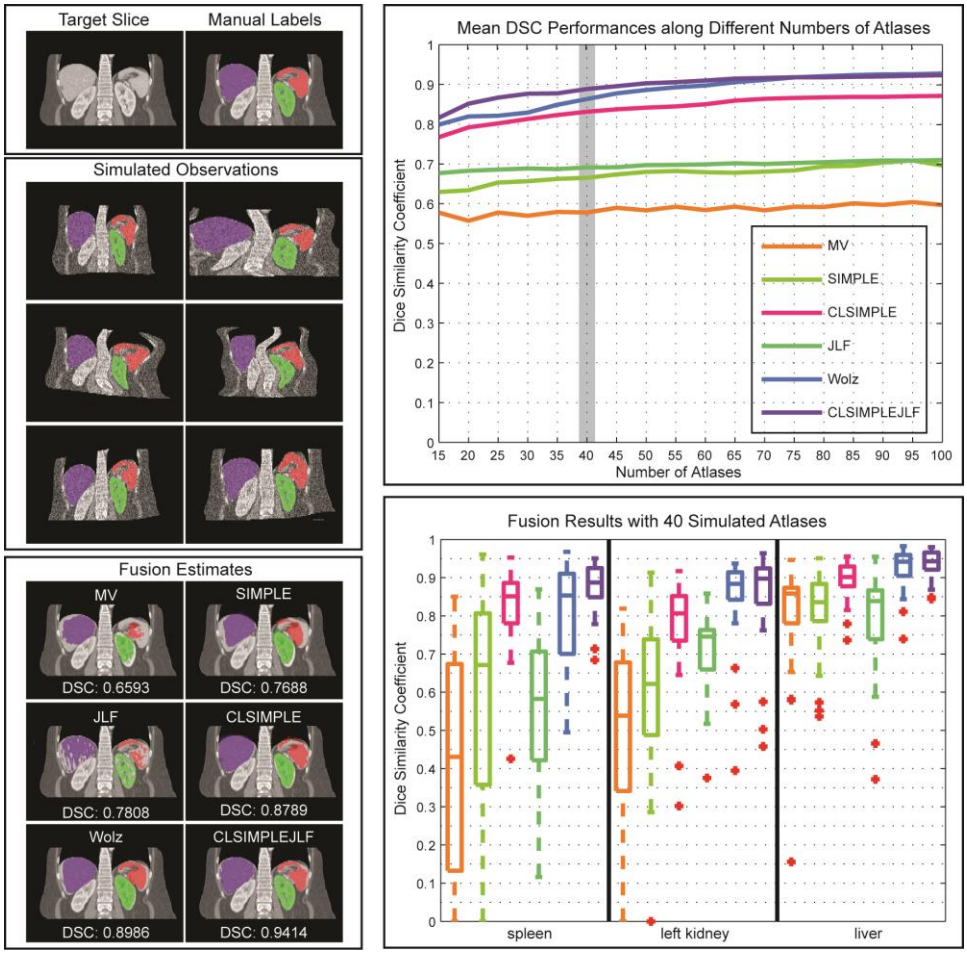

Figure VIII.4. (top left) Target slices and the associated manual labels. (middle left) Simulated observations drawn from an individual target slice with a randomly generated transformation model. (top right) The mean DSC (over 40 target slices and three organs) values evaluated for six label fusion approaches using different numbers (from 15 to 100) of atlases. (bottom right) Organ-wise DSC performances for the fusion results using 40 simulated atlases. (bottom left) Fusion estimates using 40 simulated atlases overlaid on a representative target slice, and annotated with the mean DSC value over the organs.

deformation effect on each dimension were drawn from a zero-mean Gaussian distribution with standard deviations of 3 mm. Voxel-wise Gaussian random noise (with a standard deviation of 100 Hounsfield units) was added to the simulated intensity images.

Six MAS methods, i.e., MV, SIMPLE, JLF, CLSIMPLE, CLSIMPLEJLF, and the Wolz approach were applied to 40 target slices using different numbers of atlases (from 15 to 100, with a step

size of 5), and then evaluated based on the DSC values of spleen, left kidney, and liver. Note that (1) CLSIMPLE used MV, while CLSIMPLEJLF used JLF for label fusion after atlas selection; (2) We did not append GC to smooth the results of CLSIMPLEJLF and the Wolz approach since no surface distance error was assessed in this simulation. (3) The Wolz approach here used the simulated atlases for all three stages of subject-specific atlas construction given no other intermediate registered atlases.

3.3.2. Results

Under the tests using various numbers of atlases, CLSIMPLE, CLSIMPLEJLF, and the Wolz approach demonstrate consistently and substantially more accurate segmentations than MV, SIMPLE, and JLF. CLSIMPLEJLF and the Wolz approach yield similar accuracies when using larger than 70 atlases (p-value < 0.05, paired t-test), while CLSIMPLEJLF performs better with less atlases available.

Using 40 atlases, the spread of DSC values demonstrate significant improvement by incorporating context learning. CLSIMPLE achieves a median DSC improvement of 0.26 and 0.15 over MV and SIMPLE, respectively, while CLSIMPLEJLF outperforms JLF by 0.19. CLSIMPLEJLF also provides the least range of DSC values, and thus indicates its robustness to the outliers. A representative fusion result represents that CLSIMPLEJLF accurately captures the shape, location, and orientation of the spleen, left kidney, and liver.

3.4. Volumetric Multi-Organ Multi-Atlas Segmentation

3.4.1. Experimental Setup

Ten of the 100 subjects were randomly selected as training datasets for context learning (these ten subjects happen to be all within the 75 liver cancer datasets), thus the segmentations were validated on the remaining 90 subjects. From the same cohort, forty subjects were randomly selected (independent from the ten selections for context learning) as the atlases for validating five MAS approaches, including MV, SIMPLE, JLF, the Wolz approach, and our proposed method (CLSIMPLEJLFGC), on the

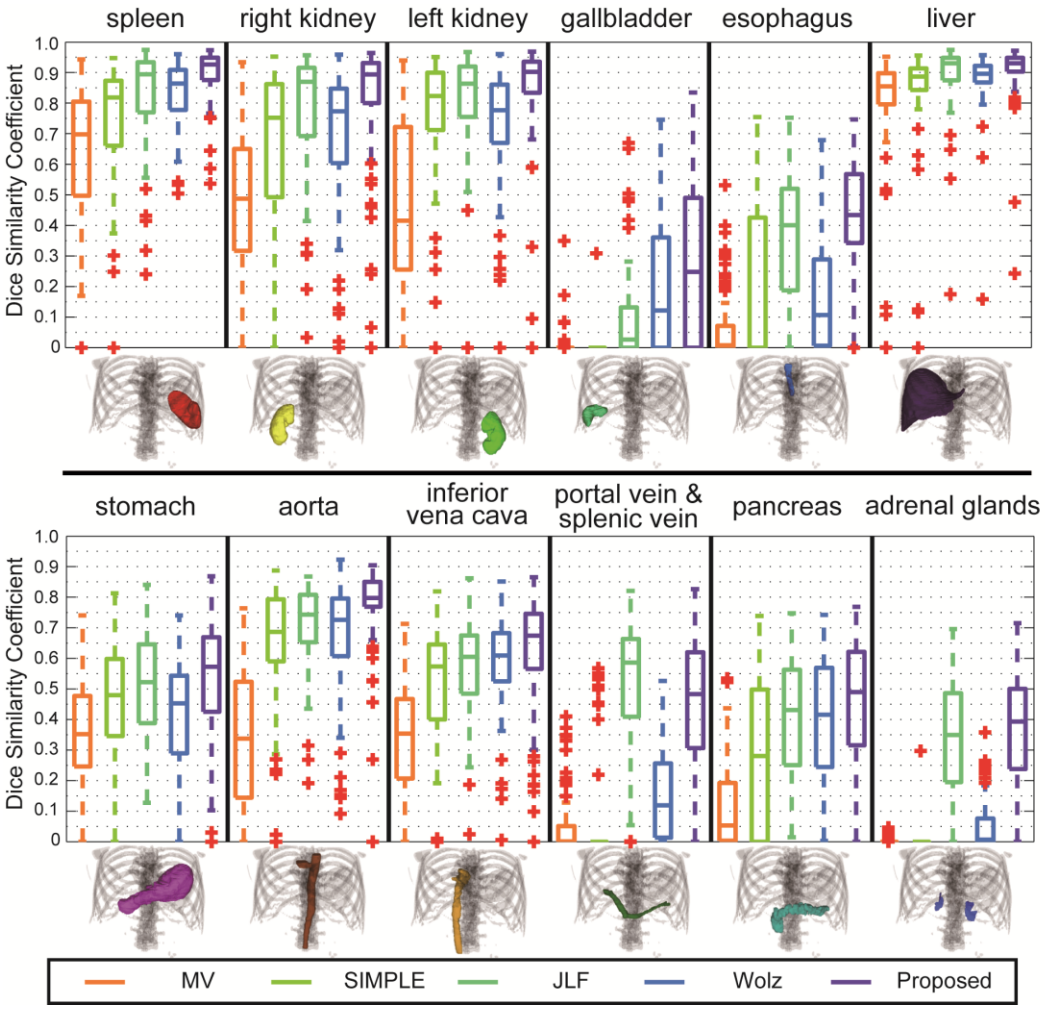

Figure VIII.5. Boxplot comparison among five tested methods for 12 organs.

segmentation of twelve abdominal organs against the manual labels using DSC, mean surface distance (MSD), and Hausdorff distance (HD).

The five approaches shared a common multi-stage registration procedure for each of the 90 target images (excluding the 10 context learning), where all atlases (except the target if it was selected in the set of atlases) were aligned to the target in the order of rigid, affine and a multi-level non-rigid registration using free-form deformations with B-spline control point spacings of 20, 10, and 5mm [95]. In summary, (1) the 10 context learning datasets were never used as targets (but they were allowed to be atlases), and (2) an atlas image was never used as its own target. Randomization of selecting context learning datasets and atlas images was performed to maximize the available data subject to these constraints.

3.4.2. Results

Compared to the other MAS approaches, the proposed method presents consistently improved segmentation in DSC on 11 of 12 organs of interest (**Figure VIII.5**, **Table VIII.1**). Based on the mean

Table VIII.1. Quantitative evaluation for five tested methods using dice similarity coefficient (mean ± std.).

	MV	SIMPLE	JLF	Wolz	Proposed
Spleen	0.63 ± 0.24	0.73 ± 0.22	0.84 ± 0.15	0.83 ± 0.10	**0.90 ± 0.08****
R. Kidney	0.47 ± 0.26	0.65 ± 0.27	0.79 ± 0.19	0.70 ± 0.24	**0.81 ± 0.20**
L. Kidney	0.46 ± 0.27	0.74 ± 0.25	0.81 ± 0.17	0.72 ± 0.21	**0.84 ± 0.20**
Gallbladder	0.01 ± 0.04	0.00 ± 0.03	0.09 ± 0.15	0.19 ± 0.21	**0.27 ± 0.26***
Esophagus	0.07 ± 0.11	0.20 ± 0.25	0.37 ± 0.21	0.18 ± 0.19	**0.43 ± 0.18***
Liver	0.79 ± 0.20	0.84 ± 0.18	0.89 ± 0.11	0.88 ± 0.09	**0.91 ± 0.09**
Stomach	0.34 ± 0.18	0.46 ± 0.19	0.51 ± 0.17	0.41 ± 0.19	**0.55 ± 0.18**
Aorta	0.34 ± 0.22	0.64 ± 0.22	0.72 ± 0.13	0.67 ± 0.18	**0.77 ± 0.13***
IVC	0.33 ± 0.18	0.50 ± 0.21	0.57 ± 0.15	0.58 ± 0.15	**0.62 ± 0.19**
PV & SV	0.05 ± 0.10	0.05 ± 0.15	**0.52 ± 0.20****	0.16 ± 0.16	0.45 ± 0.21
Pancreas	0.11 ± 0.13	0.27 ± 0.25	0.40 ± 0.19	0.40 ± 0.19	**0.45 ± 0.21**
A. Glands	0.00 ± 0.01	0.00 ± 0.03	0.34 ± 0.20	0.05 ± 0.08	**0.36 ± 0.19**

* indicates that the DSC value was significantly higher than the second best DSC across the methods for the organ segmentation as determined by a right-tail paired t-test with $p<0.05$. ** indicates a $p<0.01$.

Table VIII.2. Quantitative evaluation for five tested methods using mean surface distance (mean ± std.) in mm.

	MV	SIMPLE	JLF	Wolz	Proposed
Spleen	6.44 ± 4.30	4.42 ± 3.55	4.38 ± 7.44	3.06 ± 2.21	**1.75 ± 1.71****
R. Kidney	7.81 ± 6.73	5.22 ± 5.85	4.81 ± 10.38	4.80 ± 5.66	**2.99 ± 3.92****
L. Kidney	6.55 ± 4.63	2.92 ± 2.95	5.38 ± 11.12	3.85 ± 3.01	**2.00 ± 2.80****
Gallbladder	12.88 ± 8.29	N/A†	21.84 ± 29.35	**11.89 ± 10.53****	14.36 ± 20.34
Esophagus	7.59 ± 3.20	**3.73 ± 1.73**	7.61 ± 15.26	6.76 ± 3.70	4.16 ± 2.05
Liver	7.42 ± 9.21	5.03 ± 6.02	4.69 ± 7.01	4.86 ± 5.48	**3.22 ± 4.43***
Stomach	16.06 ± 6.61	10.96 ± 5.18	**8.75 ± 6.92***	16.91 ± 8.15	10.26 ± 6.36
Aorta	10.18 ± 7.43	4.26 ± 3.53	5.89 ± 12.83	4.68 ± 3.74	**3.02 ± 2.27****
IVC	7.92 ± 5.35	4.32 ± 1.82	6.36 ± 13.77	4.41 ± 2.38	**3.75 ± 1.84****
PV & SV	20.00 ± 5.54	6.37 ± 3.18	7.24 ± 11.61	17.46 ± 7.54	**5.92 ± 5.08****
Pancreas	16.08 ± 8.81	6.51 ± 3.96	8.24 ± 12.52	7.82 ± 4.75	**5.47 ± 3.51****
A. Glands	19.88 ± 6.43	N/A†	7.75 ± 15.12	13.30 ± 8.71	**4.06 ± 3.56***

† N/A was assigned when the segmentations were empty, and the MSD could not be computed for over 75 subjects (at least 15 subjects were not empty);

* indicates that the MSD value was significantly lower than the second lowest MSD across the methods for the organ segmentation as determined by a left-tail paired t-test with $p<0.05$. ** indicates a $p<0.01$.

DSC of each organ, a median improvement of 7.0% and 16.2% were achieved over JLF and Wolz, respectively. The segmentations of spleen, gallbladder, esophagus, and aorta using the proposed method significantly outperformed those using the other approaches.

The serpentine labels of portal vein and splenic vein are barely captured by registration (0.06 in DSC by median), thus the intermediate voting-based fusion estimates had a good chance of missing the structure entirely (zero median in DSC for MV and SIMPLE). A MV fusion (instead of JLF) of the selected atlases by SIMPLE context learning identified this structure better (0.25 in DSC by median). While with limited atlases of catastrophic registration errors, our proposed method was outperformed by JLF with all available atlases.

On the other hand, in the context of reasonably substantial registration errors for other organs, our proposed method yields segmentation with better performances in not only accuracies, but also efficiencies. With much fewer atlases (while more target-alike than average) included for label fusion, our method (1.5 hours, 10G RAM) relieved massive computational time and memory required by JLF (22 hours, 30G RAM) and Wolz approach (30 hours, 10G RAM), and thus provides more efficient abdominal segmentations. As found in our previous study [42], the MV fusion of the registered atlases with the top five DSC achieves a median DSC of 0.9 for spleen. Therefore, we considered the global non-linear selection of the atlases as a necessary procedure in addition to the locally weighted label determination for

Table VIII.3. Quantitative metrics of the proposed segmentation method.

Metrics / *Organs*	*Dice Similarity Coefficient* **Median [Min, Max]**	*Surface Distance (mm)* **Sym. HD**
Spleen	0.93 [0.54, 0.97]	17.27 ± 8.42
R. Kidney	0.89 [0.00, 0.96]	19.47 ± 11.37
L. Kidney	0.90 [0.00, 0.97]	16.13 ± 8.05
Gallbladder	0.25 [0.00, 0.84]	34.57 ± 22.87
Esophagus	0.43 [0.00, 0.75]	17.97 ± 5.46
Liver	0.93 [0.24, 0.97]	34.46 ± 15.03
Stomach	0.57 [0.00, 0.87]	49.48 ± 18.91
Aorta	0.80 [0.00, 0.90]	23.23 ± 10.98
IVC	0.67 [0.00, 0.87]	19.89 ± 5.60
PV & SV	0.48 [0.00, 0.83]	38.37 ± 17.18
Pancreas	0.49 [0.00, 0.77]	31.34 ± 8.92
A. Glands	0.39 [0.00, 0.72]	20.68 ± 8.68

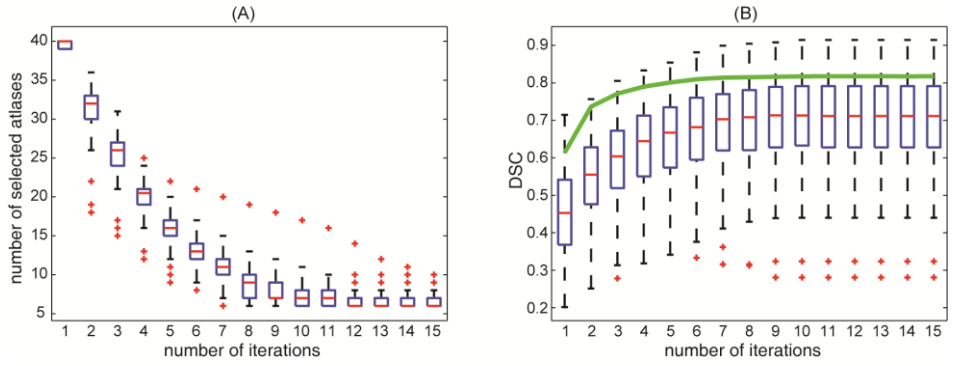

Figure VIII.6. Demonstration of the effectiveness of CLSIMPLE atlas selection for spleen segmentation on 90 subjects along number of iterations (A) number of selected atlases remaining along iterations. (B) mean DSC value of the selected atlases along iterations. Note the solid green line in (B) indicates the mean DSC of the majority vote fusion estimate using the selected atlases across all subjects.

MAS in abdomen.

With a closer look, our proposed method yielded the segmentation with at least 0.89 in DSC and less than 3.3 mm in MSD for the major organs of interest, i.e., spleen, kidneys, and liver. For other structures, the proposed method also provided successful identification over half of the subjects, even those that empirically considered difficult to capture, e.g., adrenal glands (**Tables VIII.2, VIII.3**). Qualitatively, the segmentation on a subject with median accuracy captures the organs from the perspective of both 3-D rendering and 2-D coronal slices (**Figure VIII.7**). As a side note, applying GC

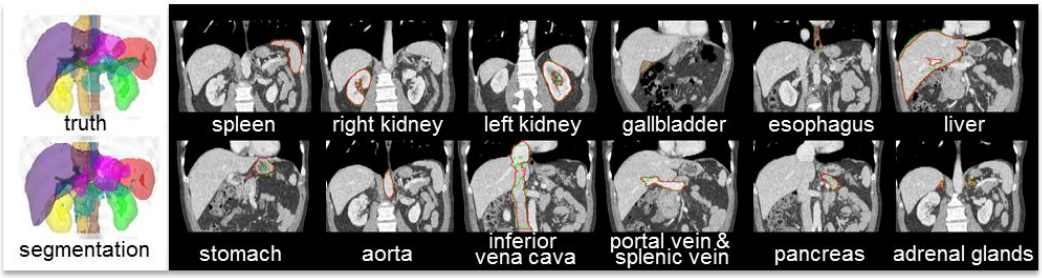

Figure VIII.7. Qualitative segmentation results on a subject with median DSC. On the left, the 3-D organ labels are rendered for the true segmentation, and the proposed segmentation. On the right, the truth (red) and the proposed segmentation (green) for each organ of interest are demonstrated on a representative coronal slice.

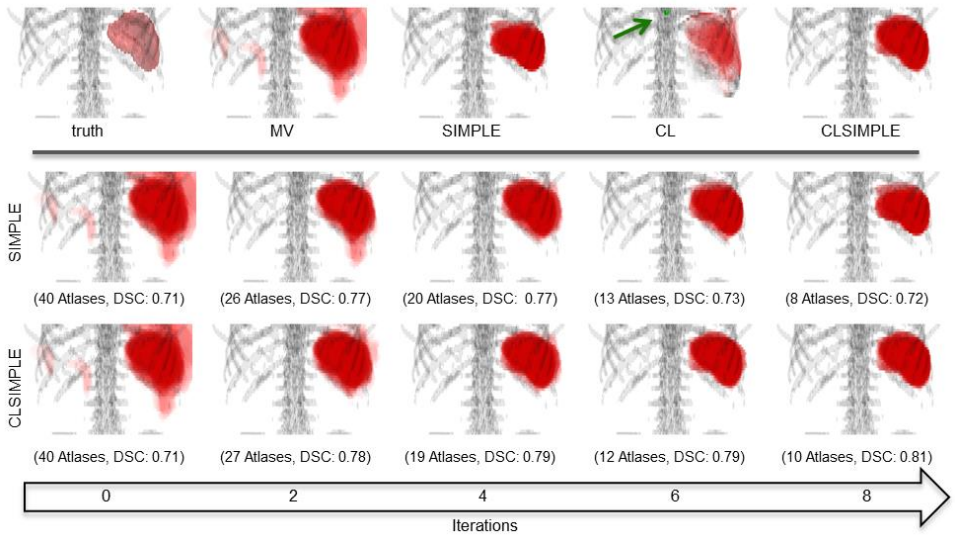

Figure VIII.8. (upper pane): The ground truth surface rendering and the probability volume rendering of different methods for spleen segmentation. Note that the transparencies of volume rendering were adjusted for visualization. CL indicates the posterior probability of spleen when applying the trained context learning model to the target. The green arrow points at the landmark used for deriving spatial context. (lower pane): Progressive results of SIMPLE and CLSIMPLE along iterations. Note that both methods reach the convergence within 8 iterations in this case.

for the five large organs (i.e., spleen, kidneys, liver, stomach) reduces the HD by 1.99 mm ($p < 0.001$, paired *t*-test), with similar the DSC values ($\Delta = -0.0038, p < 0.01$, paired t-test).

In a retrospective analysis, CLSIMPLE demonstrates effective atlas selection for spleen along iterations in terms of the mean DSC of the selected atlases and their MV fusion estimate (**Figure VIII.6**). Comparing to the original SIMPLE on an example with median accuracy, CLSIMPLE keeps adjusting the atlas selection with learned context on the target image as opposed to yielding progressively biased intermediate fusion estimate if only the registered labels were used (**Figure VIII.8**).

In a further test on the parameter sensitivity of the proposed method, ten subjects were randomly selected from the 90 subjects used for validation. The impact of using different values of three parameters, i.e., (1) number of atlases minimally allowed in CLSIMPLE, (2) patch radius in JLF, and (3)

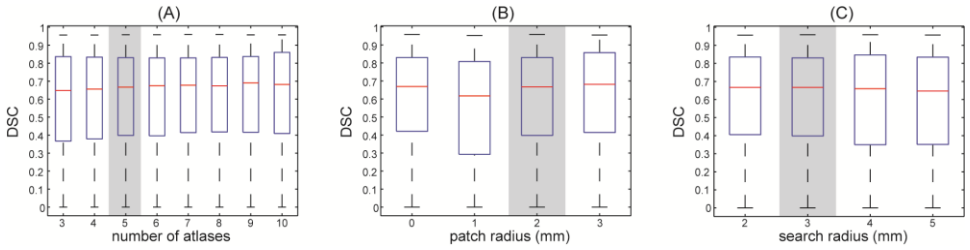

Figure VIII.9. Illustration of parameters sensitivity of the proposed method. The overall DSC values (including all twelve organs on ten subjects) are evaluated on different values of (A) number of atlases minimally allowed in CLSIMPLE; (B) patch radius in JLF; and (3) search radius in JLF. Note when testing on one parameter, the other two keep as the values the gray backgrounds; these values are also used for the segmentation of 90 subjects.

search radius in JLF, on the overall performances of the proposed method is shown in **Figure VIII.9**. Comparing to the parameters values chosen for the validation of 90 subjects, potential improvement were observed with more atlases (9 atlases, $\Delta = 0.0231, p < 0.05$, paired left-tail Wilcoxon signed rank test), and larger patch radius (3x3x3 voxels, $\Delta = 0.0143, p < 0.005$, paired left-tail Wilcoxon signed rank test).

4. Discussion

The proposed method provides a fully automated approach to segment twelve abdominal organs on clinically acquired CT. The SIMPLE context learning reduces the impact of the vastly problematic registrations with appropriate atlas selection considering exogenous contexts in addition to intermediate fusion estimate, and thus enables more efficient abdominal segmentations. We note that proposed generative model naturally leads to an iterative atlas selection, which differs from the STEPS approach [141] that first locally ranks atlases, and uses the top local atlases for statistical fusion.

MAS has been widely used for segmenting brain structures; commonly accepted optimal number of the included atlases is approximately 10 to 15. While the registration errors for brains are well constrained within the cranial vault, the registrations for abdomens, on the other hand, have much more chances to fail in terms of both global alignment and internal correspondence. Thus an atlas selection

procedure along with more included atlas images becomes essential to MAS for abdominal organs, where the effectiveness of atlas selection determines the segmentation robustness. It can be also expected that this atlas selection procedure can be beneficial for brain segmentation among subjects with substantial aging and pathological variations. The Wolz approach selects/weights atlases based on the similarity between the target and atlases in a hierarchical manner, which turns out to be more effective for the homogeneous datasets in the simulation than it is for the clinically acquired datasets. We posit that the inconsistent performances of the Wolz approach lie in the non-robust efficacy of similarity measure as discussed in the original SIMPLE literature [127]. Using the SIMPLE context learning framework, our proposed method yields consistently good performances in both datasets.

Some specific approaches for single organ segmentation, e.g., liver [198] and pancreas [19], can provide higher performances, while our efforts in this study focus on the development of a generic approach for multiple organ segmentation. In addition, provided with adequate number (>20) of labeled atlases, we expect that our proposed method can be adapted to other thoracic (e.g., lungs), abdominal (e.g., psoas muscles), and pelvic (e.g., prostate) organs on CT, where the organs to segment have (1) consistent intensity-based and spatial appearance, (2) high contrast to the surrounding tissues, and (3) reasonable amount of overlap between the registered atlases and the target. Much caution should be taken when these three conditions are not satisfied. For example, intensity normalization would be required for applications on MR images, texture-based features can be included when structures with similar intensities but distinguishable textures are close to each other, pre-localization would be necessary for tiny, thin, and/or irregular structures so that registrations errors can be constrained within the region of interest. Our future work will focus on the cases above to further improve the segmentation performances, and enhance the generalization of the method.

The estimated segmentations could be used in large-scale trials to provide abdominal surgical navigation, organ-wise biomarker derivation, or volumetric screening. The method also enables explorative studies on the correlation the structural organ metrics with surgical/physiological conditions. We note that some organs (e.g., gallbladder, portal and splenic vein, adrenal glands) have low DSC and/or

high MSD values despite the proposed method presents better segmentation over other tested MAS methods; their practice use can be limited. To our best knowledge, fully-automatic segmentation of these structures are essentially atlas-based [18, 199]. Although no ideal result has been accomplished so far, atlas-to-target registration remains the most effective approach to roughly capture these structures. Thus we present the segmentation performance for all twelve organs as a benchmark for further development. Other types of segmentation approaches, e.g., geodesic active contours [51], graph cut [101], and statistical shape models [110], are sensitive to the surrounding environment; they are often incorporated with the atlas-based framework to provide complementary information and refine the results [18, 27, 196]. Some semi-automatic approaches [200] demonstrate the potential for fundamentally better results with the requirement of manual organ identification. MAS approach performs well on automatically identifying/localizing these organs, and thus can be used as an initialization for those semi-automatic methods, and make the whole process free from manual intervention.

CHAPTER IX

IMPROVING SPLEEN VOLUME ESTIMATION VIA COMPUTER ASSISTED SEGMENTATION ON CLINICALLY ACQUIRED CT SCANS

1. Introduction

The clinical promise of computer-assisted content labeling lies in its potential to promote the extraction of quantitative morphometric information from imaging scans while minimizing time and resource requirements. The demand for such quantitative information is closely linked with the ascendency of evidence-based medicine, which depends heavily on the statistical correlation of quantitative results from different clinical datasets [201]. In time, if computer-assisted image segmentation can enable extraction of morphometric information with sufficient accuracy and time efficiency, we may advance to the point of performing high-throughput "big data" analyses of large imaging databases to discover clinically relevant associations between imaging-based phenotypic markers and other clinical endpoints, much as this approach is currently applied to high-throughput genome-wide analyses in search of novel clinically-relevant genetic biomarkers [202-205].

Among a wide range of potentially extractable morphometric biomarkers from imaging scans, lesion and organ size stand out as important targets due to the historical use of lesion and organ size information as a marker of disease presence, severity, and response to treatment [206-209]. Splenic volume is an intriguing biomarker on which to test computer-assisted segmentation techniques because of its intersection with a wide array of disease states and because of the special methodologic challenges associated with segmenting this particular organ. Quantitative estimates of spleen size have been of clinical interest for decades [210], but computer-assisted labeling of the spleen has been difficult due to wide variation between subjects in splenic size, shape, and geometric orientation within the abdomen.

Due to these difficulties, multi-atlas approaches have been pursued for identifying and labeling the spleen [20]. Briefly, this approach involves registering multiple previously labeled CT scans (atlases) to the target scan, selecting the reference scans from the atlases that are most similar to the target, and then using a statistical algorithm to combine the atlas labels from the reference scans to the target. This approach is promising for labeling anatomical structures in the abdomen because (a) it can be adapted to structures with variable image intensity and appearance, (b) it is robust to different scanning parameter settings, and usually free from interactive initialization (e.g., manual positioning of region of interest), and (c) it can be augmented by transferring newly segmented scans into existing atlases, thus allowing the algorithm to become "smarter" over time [78].

The purpose of this study was to evaluate the accuracy and time efficiency of the multi-atlas segmentation technique for estimating spleen volumes on clinically-acquired CT scans. Since preliminary work suggested that multi-atlas labeling performs exceptionally well in a majority of scans but poorly in a small subset of outliers, we investigated two "pipelines" for using multi-atlas labeling to generate splenic volumes: one pipeline consisted of computer-assisted multi-atlas segmentation on all scans, while another pipeline consisted of computer-assisted segmentation supplemented by manual segmentation for scans that failed a rudimentary and high-level visual quality check (**Figure IX.1**). We also evaluated two more traditional methods for quickly estimating splenic volume: a single-dimension craniocaudal measurement and a three-dimensional splenic index, both transformed into volume estimates by linear regression equations found in the literature [211]. We compared the accuracy of these four approaches for estimating splenic volume against ground truth values established by manual segmentation. Furthermore, since intra-subject changes in splenic volume over time may be a clinically relevant biomarker in certain settings [212, 213], we also compared the accuracy of these four approaches for estimating the change in splenic volume from prior measures for the same subject. Finally, we compared the average time for all pipelines, including manual segmentation, to generate splenic volumes.

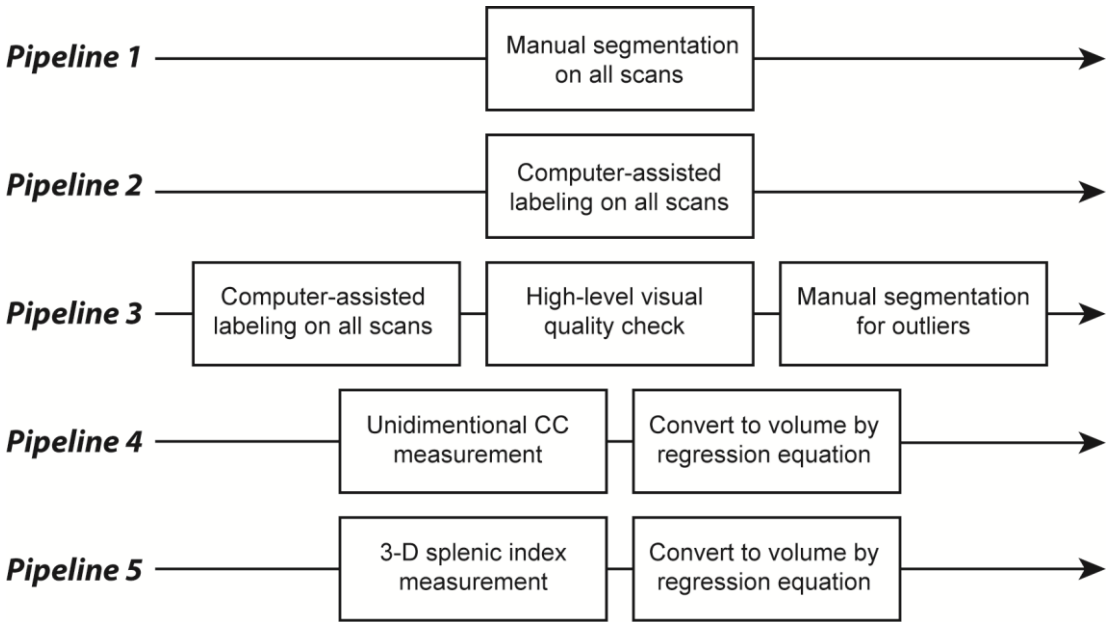

Figure IX.1. Overview of pipelines for estimating spleen volumes.

2. Methods and Materials

2.1. Data acquisition

Under institutional review board supervision and in compliance with HIPAA requirements, we obtained all contrast-enhanced CT scans performed at our institution for a recently completed phase 2 clinical trial of an investigational agent for metastatic colorectal cancer. The clinical trial cohort included 78 unique patients, all of whom had from one to eight scans during the trial, for a total of 294 abdominal CT scans. Imaging was acquired on different scanners, with varying fields of view (from 280 x 280 x 175 to 500 x 550 x 780 mm^3), in-plane resolution (from 0.54 x 0.54 to 0.98 x 0.98 mm^2), slice thickness (from 1.5 mm to 7.0 mm), and IV contrast amounts (from 70 to 120 mL iohexol 350, GE Healthcare, Waukesha, WI). Using the DCM2NII tool of the MRIcron package [214], the scans were converted from DICOM into NIFTI format for image processing. A portion of the datasets (75 out of 294) has been reported in previous technical studies [42, 44, 45]. In this manuscript, we performed our first clinical study with a larger datasets.

2.2. Manual segmentation (Pipeline 1)

An experienced research associate manually segmented the spleen on all 294 scans to obtain ground truth splenic volumes. Briefly, each scan was opened in the MIPAV software (NIH, Bethesda, MD [153]), and spleen outlines were delineated on every axial slice using an electronic pen and tablet device (Wacom Intuos, Vancouver, WA). Spleen volumes were derived by filling the regions enclosed by the spleen outlines and multiplying the counts of the labeled spleen voxels by the voxel size of the scan. A radiologist on a volumetric basis verified all splenic contours. The time from starting the labeling process to calculating the splenic volume was tracked for each scan.

Additionally, a subset of 74 scans was manually labeled by a second research associate in order to assess the inter-rater reliability of the manual segmentation process. This second research associate followed the same outline-filling and volume-measuring process as described above.

2.3. Computer-assisted segmentation (Pipeline 2)

Automated segmentation of the spleen was performed for all scans using a recently published multi-atlas label fusion technique [44]. Briefly, 65 scans were randomly selected from among the baseline scans (i.e., first time point of the clinical trial) to serve as the reference atlases. For each target image of the 294 CT scans, 65 or 64 non-rigid registrations [95] (excluding the target itself if it was within the atlases) were performed, through which the manually labeled spleens were propagated from the atlases to the target. Atlas selection followed by statistical label fusion was performed to yield the estimated spleen segmentation by combination of the registered spleen labels. Specifically, the atlases were iteratively evaluated and selected based on the volumetric overlap with the intermediate result across the remaining atlases using an Expectation-Maximization scheme, where the intermediate result was a majority-voted fusion of the remaining registered atlases adjusted by the spleen likelihood on the target scan given a Gaussian mixture model (trained with 10 of the 65 atlases) of the image contexts [42]. Joint label fusion [134] was employed on the selected atlases to yield the final segmentation; the voxel-wise decisions were based on a weighted voting of the atlases, where the weights were derived on each voxel for each atlas to

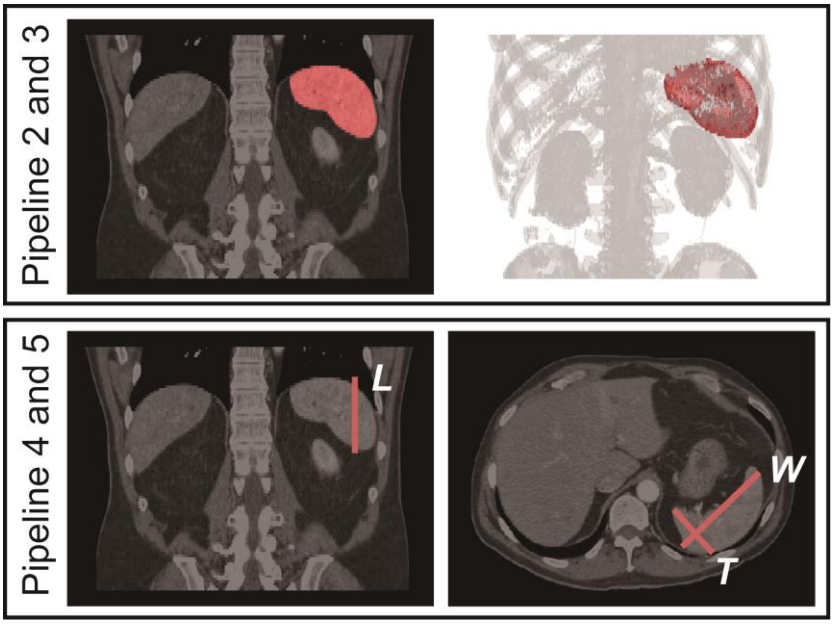

Figure IX.2. Illustration of the required measurements from different pipelines for estimating the spleen volume. Pipeline 2 and 3 extract the whole spleen volume, while pipelines 4 and 5 measure splenic diameters along different axes.

minimize the correlated errors across the atlas images in local neighborhoods. The resultant labeled regions (**Figure IX.2**) were then used to calculate the estimated splenic volume by multiplying the counts of the labeled spleen voxels by the voxel size of the scan. The total time of manual interaction for running the spleen segmentation and calculating the splenic volume for each scan was tracked.

2.4. Computer-assisted segmentation with manual labeling for outliers (Pipeline 3)

Pipeline 3 simulated a workflow in which manual segmentation was used only for those computer-segmented scans that failed a rudimentary visual quality check. The quality check, performed on all computer-segmented scans, consisted of overlaying the spleen segmentation result on single axial, coronal, and sagittal plane CT slices through the middle of the spleen (as shown in **Figure IX.3**). Based on qualitative judgment, a scan failed the quality check if its segmentation on at least one slice appeared to capture less than 50% of the true underlying structure.).

If a scan passed the quality check, the computer-assisted splenic volume was accepted. If a scan failed the quality check, the splenic volume from manual segmentation (Pipeline 1) was used. Time per

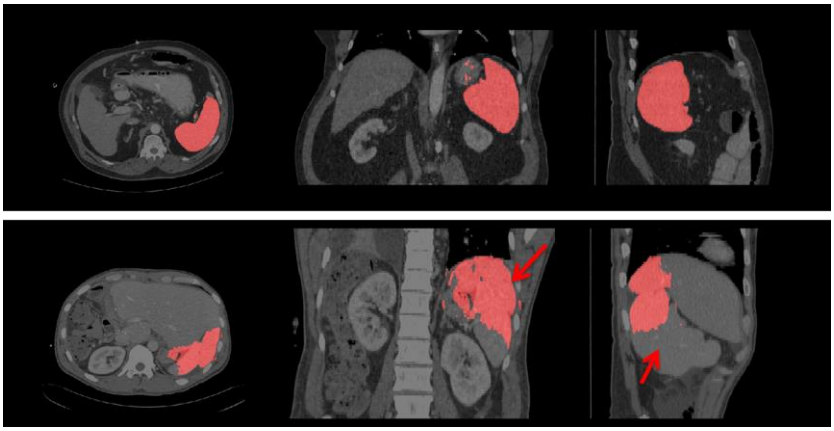

Figure IX.3. Quality assurance of the computer assisted segmentation in Pipeline 3 was performed by overlaying the spleen segmentation result on single axial, coronal, and sagittal CT slices through the middle of the spleen. Upper row: a successful case where the automated labels were used. Lower row: a failure case where manual correction was required.

scan for this pipeline was computed as the average time required to perform the quality check plus the average time to segment the images and generate a splenic volume (including all required manual interactions).

2.5. Unidimensional and splenic index measurements (Pipelines 4 and 5)

An experienced research associate extracted linear measurements of spleen size on all scans following the earlier work of Bezerra et al. [211], who compared different linear spleen measurements and indices and presented transfer equations correlating these measurements with splenic volumes. In Pipeline 4, a unidimensional measurement of craniocaudal spleen length (L) was extracted as described in [211] by multiplying the number of axial sections in which the spleen was visualized by the thickness of the sections (**Figure IX.2**); spleen length was then converted into a splenic volume using the linear regression equation L = (0.0126 x Volume) + 5.8006, again from [211]. In Pipeline 5, a three-dimensional splenic index (SI) was extracted as described in [211] by multiplying craniocaudal spleen length (L) by the maximum axial plane width (W) on any section (typically in oblique sagittal plane, see **Figure IX.2**) by the maximum axial plane thickness (T) on any section (perpendicular to W, see **Figure IX.2**); the resultant splenic index (SI = L x W x T) was then converted into a splenic volume using the linear

regression equation SI = (2.7036 x Volume) + 0.4776, again from [211]. The times to generate splenic volumes were tracked separately for Pipelines 4 and 5.

2.6. Statistical analysis

The following set of metrics were used to evaluate segmentation accuracy: Dice similarity coefficient (DSC) [146], Pearson correlation, R-squared, and absolute and percent deviation of splenic volumes from ground truth. First, in order to assess the inter-rater reliability of the manual segmentation process in Pipeline 1, these accuracy metrics were computed for the subset of 74 scans in Pipeline 1 that underwent segmentation by two independent research associates. Then, to evaluate the ability of Pipelines 2-5 to generate point estimates of splenic volume, the accuracy measures were computed for each Pipeline using the complete set of 274 manually segmented scans (Pipeline 1) as the ground truth. To evaluate the ability of Pipelines 2-5 to estimate changes in splenic volume between successive measures for an individual subject, the accuracy metrics were again computed for Pipelines 2-5 using the information from Pipeline 1 as the ground truth. These computations were performed both for change in splenic volume from baseline (i.e., first time point of the clinical trial) and for change from the most recent prior scan.

One-tail Wilcoxon signed rank tests were performed to compare the accuracy measures (DSC and absolute deviation) if applicable. The significance tests were performed again for the time cost between Pipelines 3-5. Bland-Altman plots were constructed for each of Pipelines 2-5 to depict the agreement against Pipeline 1 on the point estimates of splenic volume (**Figure IX.4**).

3. Results

3.1. Inter-rater reliability of manual segmentation

Metrics for Pipeline 1 inter-rater reliability are presented in **Table IX.1**. The manual spleen segmentation demonstrated high reproducibility with an average DSC of 0.96 between labels from two

individual raters. The extract splenic volume sizes were also highly correlated in terms of the Pearson correlation and R-squared.

3.2. Point estimates of splenic volume

For Pipeline 2, the computer-assisted segmentation achieved a DSC of 0.90 ± 0.11 against the manual segmentation. 20 out of 294 scans were below 0.70 in DSC. When performing Pipeline 3, 22 out of 294 scans failed the visual quality check and required re-labeling; this approximately matched the number of scans with low DSC values. With outlier correction, Pipeline 3 achieved a DSC of 0.93 ± 0.07, and substantially better agreement in spleen volume measurement than Pipeline 2 in terms of Pearson

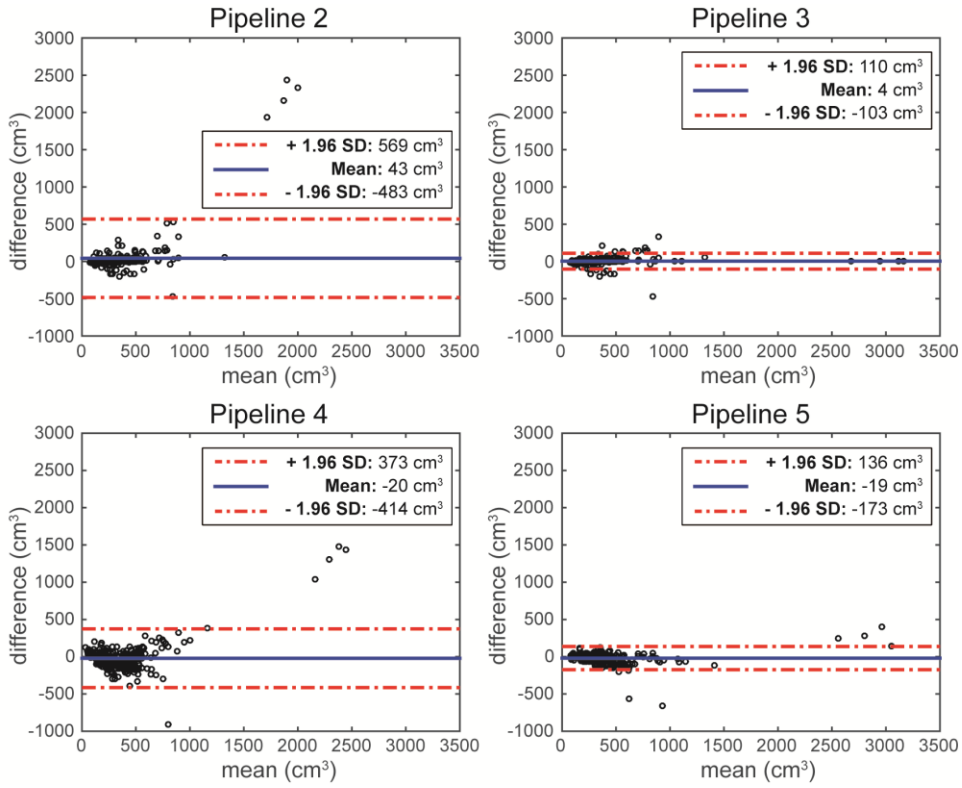

Figure IX.4. Bland-Altman plots for different spleen volume estimation methods using pipelines 2-5. On each plot, the horizontal axis represents the mean volume between the ground truth and the estimation, while the vertical axis indicates the difference in volume from the ground truth to the estimation. The mean in difference, and a confidence interval of ± 1.96 standard deviation (SD) are highlighted.

correlation (0.98 vs. 0.72) and R-Squared (0.98 vs. 0.51).

Splenic volume estimation using 3-D splenic index (Pipeline 5) performed better than 1-D splenic length (Pipeline 4). Close values of Pearson correlation and R-squared were observed between Pipelines 3 and 5, although Pipeline 3 yielded a significantly smaller ($p < 0.05$) absolute deviation of estimated volume from ground truth. Among Pipelines 2-5, Pipeline 3 provided the best agreement with manual

Table IX.1. Accuracy and Time Results for Each Pipeline.

Result	Manual segmentation (Pipeline 1)[a]	Computer segmentation for all scans (Pipeline 2)	Computer segmentation with manual segmentation of outliers (Pipeline 3)	1-D length (Pipeline 4)	3-D splenic index (Pipeline 5)
Accuracy (point estimates)					
Dice similarity coefficient	0.96 ± 0.01	0.90 ± 0.11	0.93 ± 0.07**	N/A	N/A
Pearson correlation	0.9997	0.7151	0.9888	0.8613	0.9765
R-squared	0.9993	0.5114	0.9778	0.7435	0.9535
Absolute deviation of volume from ground truth (cm^3)	7.25 ± 7.29	64.66 ± 263.93	23.70 ± 48.98**	111.02 ± 168.48	46.92 ± 66.37
Accuracy (change from baseline)					
Pearson coefficient	N/A	0.5556	0.8741	0.4839	0.8178
R-squared	N/A	0.3087	0.7641	0.2717	0.7437
Absolute deviation of volume change from ground truth (cm^3)	N/A	46.70 ± 84.86	28.24 ± 49.55**	81.62 ± 107.82	38.11 ± 68.01
Accuracy (change from most recent prior)					
Pearson coefficient	N/A	0.6532	0.8094	0.4825	0.7694
R-squared	N/A	0.4267	0.6551	0.2590	0.6597
Absolute deviation of volume change from ground truth (cm^3)	N/A	38.04 ± 67.07	26.33 ± 52.12**	64.28 ± 81.64	33.95 ± 59.34
Time cost					
Manual interaction time (averaged per scan)	11 min	5 sec	1 min*	1 min 5 sec	1min 30 sec

[a] measured between the labels from two independent raters on a subset of 74 scans

* Dominant at a $p<0.05$ using one-tail Wilcoxon signed rank test between Pipelines 3-5

** Dominant at a $p<0.05$ using one-tail Wilcoxon signed rank test (comparing to the second best)

segmentation as depicted by the Bland-Altman plots (**Figure IX.4**).

3.3. Estimates of change in splenic volume

For both change from baseline and change from most recent scan, Pipeline 3 again produced the most accurate results, with Pipeline 5 the second most accurate. In general, the estimates for change in splenic volume were not as accurate as the single point estimates for splenic volume.

3.4. Time costs

Table IX.1 reports the manual interaction time for each method. Pipeline 3 included 50 seconds for manual labeling, 5 seconds for running segmentation program, and 5 seconds for quality assurance. For Pipelines 4 and 5, the time include an average of 39 seconds overhead time to search for the scan (by medical record number and scan date), load the scan, cache the images for analysis, and close the scan.

4. Discussion

This study demonstrates the feasibility of using computer-automated segmentation to extract spleen volumes on large CT datasets. Pipeline 3 – i.e., computer segmentation for a majority of scans and manual segmentation for a minority of easily recognized outliers – was dominant among Pipelines 2-5 in terms of both accuracy and time cost, achieving a Pearson correlation coefficient of 0.99 and an average absolute deviation of 23.70 cm^3 from ground truth spleen volumes at a time cost of only 1 minute per scan. This approach was superior to computer segmentation (Pearson correlation 0.72, absolute deviation 64.66 cm^3, time cost 5 seconds per scan), unidimensional measurement (Pearson correlation 0.86, absolute deviation 111.02 cm^3, time cost 1 minute 5 seconds per scan), and 3-D splenic index measurement (Pearson correlation 0.98, absolute deviation 46.92 cm^3, time cost 1 minute 30 seconds per scan). These results show that a hybrid approach combining fully automated processing with manual correction can enable generation of accurate spleen volumes with reasonable time efficiency.

In reviewing the computer-automated segmentations, we found that 90% of the spleen volumes from Pipeline 2 were highly accurate, but about 10% of outliers were highly inaccurate. Typical outliers

include spleens with large volumes (at least twice as the average), and those abut other organs (liver and left kidney). These outlier segmentations were easily spotted with a rudimentary visual quality check, however, and when these outliers were manually segmented, the average time for manual interaction was still less than Pipelines 4 and 5.

The presented computer-assisted segmentation approach achieved spleen volume accuracies comparable to other recent related works. Wolz et al. used a hierarchical scheme of multi-atlas segmentation and achieved a DSC of 0.92 for spleen [28]. Linguraru et al. presented a spleen segmentation tool with an average DSC of 0.95 using probabilistic atlases followed by the correction of shape, location, and contrast enhancement [20]. Okada et al. modeled the inter-organ relationship of shape and location in the abdomen, and yield the spleen segmentation with a DSC of 0.93 [181]. We note that unlike other prior attempts, our dataset was comprised of clinically acquired CT scans at conventional doses.

The presented multi-atlas label fusion approach may be further improved from two perspectives. First, pre-localizing the spleen region of interest (ROI) can largely decrease the registration failures and thus enhance the segmentation performances; this enhanced version has been implemented where better accuracies were observed [45]. Second, a prospective feedback loop can be constructed where datasets failing the quality check would be appended to the atlas set. For the multi-atlas label fusion technique, it is critical to select atlases similar to the scan, especially when there are large shape and context variations for the structure of interest. A growing atlas set can benefit spleen segmentation on later datasets by providing larger coverage of spleen variations to capture the extreme cases. For example, in this study, most of the scans requiring re-labeling had a large spleen volume, which was beyond the capability of the current atlas set. It can be expected that an augmented atlas set including these scans could capture large spleens better in the future.

We see a variety of potential clinical applications for this technique. Spleen volumes are relevant biomarkers for the presence of many disease processes (e.g., small spleen in sickle cell disease; large spleen in portal hypertension, certain hematologic malignancies, infiltrative processes such as Gaucher's

and glyocogen storage diseases, and infections such as HIV and infectious mononucleosis). Tracking spleen volume changes over time has a potential clinical role for monitoring disease progression, as a prognostic indicator for disease complications (e.g., esophageal varices), and as a biomarker for treatment response (e.g., response to therapy in myelofibrosis [213, 215]). While spleen volumes are currently most often described on clinical CT reports in qualitative terms, quantification offers the potential for increased diagnostic accuracy, decreased subjectivity of interpretation, and higher sensitivity for subtle and potentially clinically relevant changes. Computer-assisted segmentation applied to large population datasets has the potential to improve on currently accepted reference values for abnormal spleen volumes adjusted for covariates including age, gender, and body mass index. Finally, our proposed multi-atlas fusion approach provides information not just on splenic volume, but also on geometric features associated with the shape of the organ itself, which may provide for interesting and relevant correlations when applied to high-throughput "big data" phenotype-clinical analyses.

In summary, we presented a hybrid computer-assisted segmentation method with manual segmentation of outliers that provides accurate estimation of spleen volumes at a reasonable time cost and opens the opportunity for a variety of potential clinical applications. We expect our multi-atlas fusion algorithm to improve – and thus the need for manual segmentation of outliers to decrease –with increasing size of the reference atlas set. We envision a future in which this technique may be incorporated into PACS software for routine reporting of quantitative morphometric information, not only for spleen but for other organs and for discrete lesions as well. The path for clinical translation likely involves incorporation first into clinical trials and specialized core laboratories, and subsequently into broad clinical use as time efficiencies improve.

CHAPTER X

EVALUATION OF SIX REGISTRATION METHODS FOR THE HUMAN ABDOMEN ON CLINICALLY ACQUIRED CT

1. Introduction

The human abdomen is an essential, yet complex body space. Bounded by the diaphragm superiorly and pelvis inferiorly, supported by spinal vertebrae, and protected by muscular abdominal wall, the abdomen contains organs involved with blood reservation, detoxification, urination, endocrine function, and digestion, and includes many important arteries and veins. Computed tomography (CT) scans are routinely obtained for the diagnosis and prognosis of abdomen-related disease; yet no specific image registration tools for the abdomen have been developed.

General-purpose registration tools (initially designed for volumetric brain registration) are being applied to abdominal CT scans [28, 44] On abdominal CT, inter-subject variability (e.g., age, gender, stature, normal anatomical variants, and disease status) can be observed in terms of the size, shape, and appearance of each organ. Soft anatomy deformation further complicates the registration by varying the inter-organ relationship, even within individuals (e.g., pose, respiratory cycle, edema, digestive status). Hence, characterization of tools specifically on abdominal structures is necessary, as opposed to relying on brain-centric reviews [93].

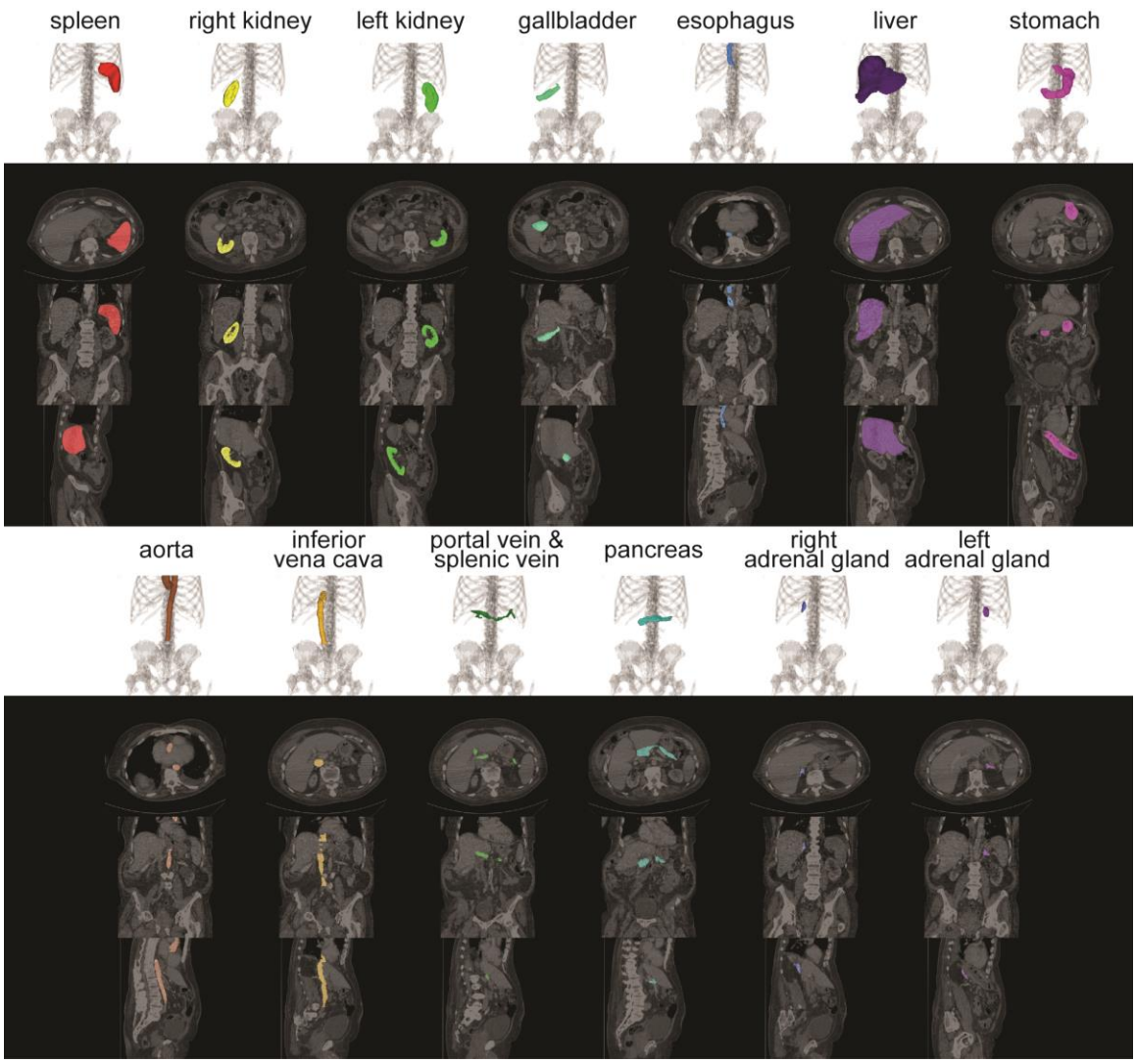

Figure X.1. Illustration of 13 organs of interest on volumetric rendering and 2-D slices of axial, coronal and sagittal orientations.

This work follows the framework of Klein et al. [93], in which 14 nonlinear registration tools and one linear registration algorithm were applied to 80 MRIs of the human brain. Manual segmentations of regions are used to assess volumetric overlap and surface-based criteria separately from the intensity-based metrics that drive registration. In related work, West et al. [216] established a platform for assessing landmark-based registrations on retrospective intermodality (MR, CT, and PET) brain images, where 12 methods were evaluated based on target registration error [217]. Murphy et al. [218] compared

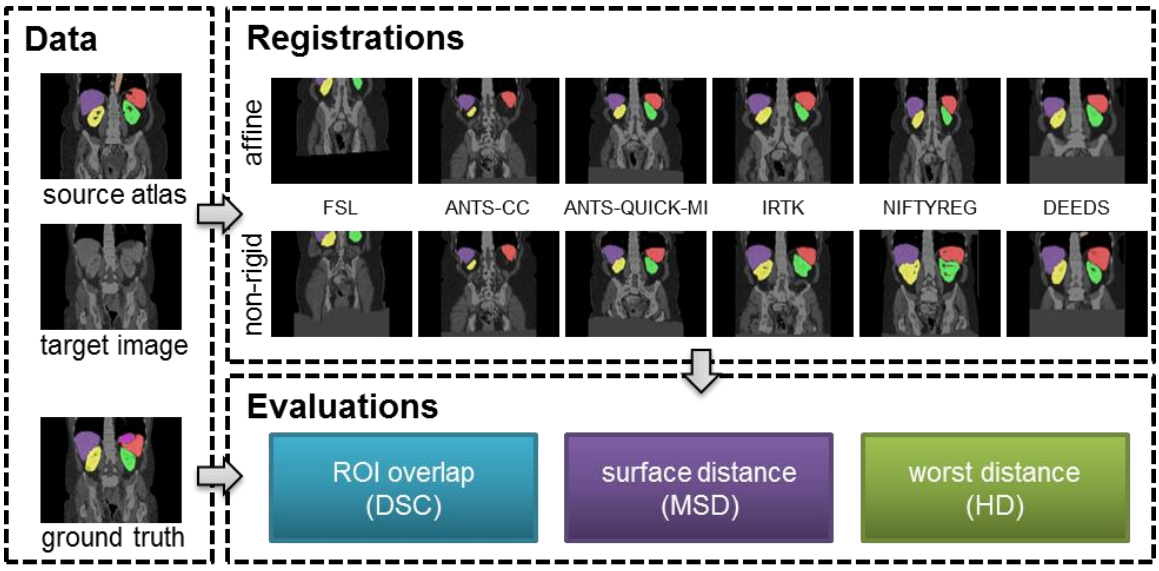

Figure X.2. Registration pipeline. Given a pair of target image and a source atlas (image and labels), an affine registration was applied followed by a non-rigid registration for each of the six evaluated registration methods. The registered labels were validated against the ground truth (manual labels) in terms of DSC, MSD, and HD.

20 registration algorithms to 30 thoracic CT pairs in the EMPIRE10 challenge by metrics specified for pulmonary area alignment and correspondence.

This work expands on [41] by including more datasets (100 vs. 20), adjusting the label sets (the previous individual label of the adrenal glands were separated into two labels: right and left), using a different registration framework (previously all non-rigid registrations were initialized by one affine registration tool), and presents more comprehensive statistical analyses (see in methods section) (**Figure X.1**). We selected 5 registration tools that have been successful in volumetric brain registrations, including FSL [219], IRTK [95], NiftyReg [188], ANTs [96], and DEEDS [220] due to their academic popularity and general availability. In total, six registration methods were evaluated with two different parameter settings for ANTs. For each registration tool, we applied affine registration followed by non-rigid registration. Registration results from both stages were evaluated based on Dice similarity coefficient (DSC [146]), mean surface distance (MSD), and Hausdorff distance (HD). We note that substantial registration errors can be expected in this study since no image registration tools are tailored

for abdomen as for the brain and thorax registrations. We also note that the efficacy of non-rigid registrations are greatly impacted by the baseline affine registrations as a lesson learned from [41], thus we modified the registration framework to use affine and non-rigid registration from the same registration tool.

2. Methods

The registration evaluation process follows the flowchart in **Figure X.2**.

2.1. Data Acquisition

Under institutional review board supervision, 100 abdominal CT scans were collected anonymously from two clinical trials. From an ongoing colorectal cancer chemotherapy trial, the baseline session of the abdominal CT scans were randomly selected from 75 metastatic liver cancer patients; the remaining 25 scans were acquired from a retrospective post-operative cohort with suspected ventral hernias. All 100 scans were captured during portal venous contrast phase with variable volume sizes (512 x 512 x 53 ~ 512 x 512 x 368) and field of views (approx. 280 x 280 x 225 mm^3 ~ 500 x 500 x 760 mm^3). The in-plane resolution varies from 0.54 x 0.54 mm^2 to 0.98 x 0.98 mm^2, while the slice thickness ranged from 1.5 mm to 7.0 mm. All image scans and their associated labels were converted to NIFTI format with the DCM2NII tool of the MRIcron package [214]. The image orientations in the NIFTI header describe the relative position of patients with respect to the scanner. Due to the inconsistencies of scanning protocols, the images were re-oriented to standard orientation with the FSL package before any further processing [219].

Thirteen abdominal organs were considered regions of interest (ROI), including spleen, right kidney, left kidney, gall bladder, esophagus, liver, stomach, aorta, inferior vena cava, portal and splenic vein, pancreas, left adrenal gland, and right adrenal gland. The organ selection was essentially based on [18]. Suggested by a radiologist, we excluded the heart for lack of full appearance in the datasets, and included the adrenal glands for clinical interests. These ROIs were manually labeled by two experienced

undergraduate students, and then verified by a radiologist on a volumetric basis using the MIPAV software [153]. A subset of 13 scans was randomly selected, and independently labeled by each of the two raters. Mean overall DSC overlap between the raters (i.e., inter-rater variability) was 0.87 ± 0.13 (0.95 ± 0.04 when considering only the spleen, kidneys, and liver).

2.2. Registration Pipeline

General-purpose registration software typically provides options and parameters for specific applications. Six registration methods from six registration tools were evaluated in this study, and indicated as FSL, ANTS-CC, ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS respectively. All registration commands evaluated in this study were verified by the developers of the corresponding registration software.

All tested methods follow a standard registration pipeline: For each image pair, source (moving / floating) and target (fixed / reference) images, the registration was driven by the similarity metrics between their intensity images. The registration was divided into two stages - affine registration that aligned the two images with co-linearity persevering transformation (translation / rotation / scaling /shearing), followed by a non-rigid registration that refined the local correspondence with deformation models. Based upon the transformation / deformation generated from the intensity-driven registration, the labels associated with the source image were propagated to the target space with nearest neighbor interpolation as the estimate of the target structures.

We briefly describe the registration setups for each method without detailed parameters.

- FSL used the FLIRT and FNIRT for affine and non-registration, respectively. The affine registration with 9 degrees of freedom (DOF) was initialized by a rigid registration. Both rigid and affine registrations constrained the search of rotations with "-nosearch".
- ANTS-CC and ANTS-QUICK-MI used different parameter settings with ANTs package. The parameters were derived from the example scripts (antsRegistrationSyN and antsRegistrationSyNQuick, respectively) in the ANTs package. ANTS-CC used cross-

correlation as the image similarity metric, while ANTS-QUICK-MI used mutual information. ANTS-QUICK-MI was specified to converge with fewer iterations than ANTS-CC, and thus noted with "QUICK". Both methods applied 5 levels of multi-resolution sampling, windowed the intensity range, started with the alignment of center of mass, initialized the affine registration with rigid registration, and used symmetric normalization (SyN) transform for the non-rigid registration. Multi-thread computing was enabled to use two CPU cores for one registration process.

- IRTK sequentially used rigid, affine, and non-rigid registrations. For all three procedures, the target padding value was set to -900 reduce the impact of the background in the CT scans (air with -1024 Hounsfield unit), 3 levels of multi-resolution sampling were applied. Assuming relatively homogenous orientations of patients' bodies in the CT scan, the options of "translation_only" and "translation_scale" were specified for the rigid and affine registration, respectively, so that only translation (and scaling for the affine registration) adjustments were allowed, and the searches over rotations were prohibited. The B-spline control spacing free-form deformation for the non-rigid registration was set to be 20, 10 and 5mm for the 3 resolution levels, respectively.
- NIFTYREG used 5 levels of multi-resolution sampling for both affine and non-rigid registrations. For the non-rigid registration based on block-matching approach and free-form deformation, an upper intensity threshold of 500 was set for both target and source image, and the maximum iteration for convergence was limited to 1000. Multi-thread computing was enabled to use two CPU cores for one registration process.
- DEEDS used 5 scale levels with grid spacing ranging from 8 to 4 voxels, displacement search radii from 6 to 2 steps with quantizations between 5 and 1 voxels. The regularization weighting was set to be 0.4. Self-similarity context descriptors [221] were derived, while their Hamming distance between images were used to guide the local displacement. All scans were resampled to an isotropic resolution of $2.2mm^3$, and

cropped to have same dimensions. The non-rigid registration is initialized using an affine registration that is based on the same similarity metric, a similar block-matching search and trimmed least squares.

2.3. Running Registrations

All registrations were run on an Oracle Grid cluster of twelve 64-bit Ubuntu 14.04LTS Linux servers. Each workstation had 12 2.8GHz cores and 48 GB RAM. Each registration was specified with the approximated maximum memory usage based on their computational complexity; multiple registrations were allocated on the memory requirement on the cluster servers, and operated in parallel. The memory specified in GB for FSL, ANTS-CC, ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS were 20, 20, 20, 10, 10, and 5. Given 100 scans, 9900 sets of output registration can be generated for each method with a leave-one-out scheme. Specifically, for each target image among the 100 scans; the remaining 99 scans were used as source images to the target image in a pair-wise manner. However, during initial running trials, we found that FSL and ANTS-CC took unreasonable amount of time to complete (> 6 h, see **Table X.1**). Therefore, these two methods are only validated on a randomly selected subset of the datasets. Specifically, 20 target images and 20 source images were randomly selected without replacement from the 100 datasets, and 400 registrations were applied from all combinations of the source-target pairs. For the other four methods, i.e., ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS all 9900 registrations were applied. In total, this study used approximately 103,800 hours of CPU time for registration.

2.4. Evaluation Metrics

DSC was used to evaluate the volumetric overlap between the estimated segmentation and the truth. Briefly, consider A as the segmentation volume, B the truth volume, and $|\cdot|$ the L^1 norm operation,

$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{10.1}$$

Surface error criteria characterize how far the surfaces of the estimated segmentation and the truth are from each other. Vertices were collected from the surfaces of both the segmentation and truth, based on which distances between the sets of vertices are measured in terms of their spatial coordinates. Let the vertices on the segmentation and truth surface be X and Y, respectively, and $d(\cdot,\cdot)$ be an indicator of distance measure. Then typically, the MSD error and HD error from the segmentation to the truth can be measured as below.

$$MSD(X,Y) = \underset{y\in Y}{avg}\ \underset{x\in X}{inf}\ d\,(X,Y) \tag{10.2}$$

$$HD(X,Y) = \underset{y\in Y}{sup}\ \underset{x\in X}{inf}\ d\,(X,Y) \tag{10.3}$$

where *sup* represents the supremum, *inf* the infinum, *avg* the average. Symmetric surfaces differences were used in this study as they better capture errors between potentially rough surfaces, i.e.,

$$MSD_{sym}(X,Y) = \frac{MSD(X,Y) + MSD(Y,X)}{2} \tag{10.4}$$

$$HD_{sym}(X,Y) = \frac{HD(X,Y) + HD(Y,X)}{2} \tag{10.5}$$

All metrics were evaluated in an organ-wise manner between the registered labels (estimated segmentation) and the manual labels (ground truth).

2.5. Statistical Analyses

For each pair of methods, permutation tests were performed to examine the statistical significance for the overall DSC and MSD across all organs. Following [93, 222], each test provided an exact p-value calculated as the percentage of *N* permutations that the absolute mean differences after permutation is larger than the original absolute mean differences between the metrics of two methods on a subset of independent registration pairs, where no overlap is allowed within the images (including both target and source images) associated with the selected registrations, and thus the correlation between registrations with shared scans was prevented. The tests were repeated M times with randomized selection of subsets, and an average p–value was obtained to indicate the significant difference between tested methods. Tests involving FSL or ANTS-CC (or both) selected subsets among the 20 target images and 20 source images

(400 registrations) that these two methods had been applied, where 10 independent registration pairs could be obtained for each subset. Tests within the other four methods, i.e., ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS selected 50 independent registration pairs among 100 images (9900 registrations). In both cases, we let $N = 1000$ for the number of permutations, and $M = 10000$ for the number of random selections of subsets.

Indifference-zone ranking considers two metrics as equal when they are within a delta of one another, where the delta characterizes the practical difference [223]. We performed two groups of indifference-zone ranking to examine the practice significances for DSC and MD in an organ-wise manner among the non-rigid registrations of the tested methods. The first group included all methods with 400 registrations, while second group had ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS evaluated with 9900 registrations. For each organ, let i and j be the row and column index of an $L \times L$ matrix (L is 6 and 4 for the first and second group, respectively), L_{ij} was assigned by the values of -1, 0, or 1, for the cases when the evaluation measure for the i^{th} method was at least delta less than, within delta of, or at least delta greater than that of the j^{th} method. The outputs were then averaged across all registrations. The delta value was specified for each organ on each subject based on the surface area of organs. The surface area of an organ label was calculated by summing up the face areas in contact with the background across the foreground voxels; it was adjusted by a constant coefficient to yield a delta value that represents the practical difference of the evaluation metric. For DSC, we used a mean delta value of 0.05 for DSC across all organs and 7 mm for MSD. A higher indifference-zone score represents a better DSC performance, while a lower score was favorable for the MSD performance.

3. Results

Registrations were successful in terms of software error codes except for 6 out of 9900 ANTS-QUICK-MI failed without producing output. The evaluated metrics of the affine and non-rigid outputs on each organ were illustrated in **Figures X.3, X.4,** and **X.5** in terms of DSC, MSD, and HD, respectively.

Regarding the overall performances across all registration methods, over half of the registrations have the DSC values lower than 0.7 for the majority of the organs. The MSD and HD boxplots clearly illustrate the overbearing amount of outliers with up to 500 mm.

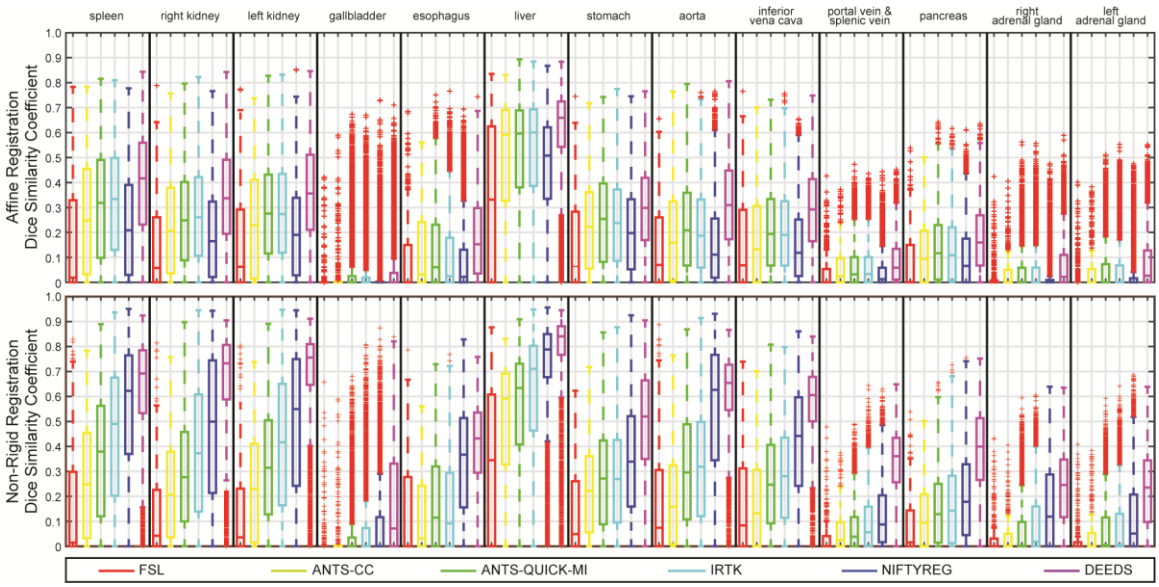

Figure X.3. Boxplot of DSC values on 13 organs for the affine and non-rigid outputs of six registration methods.

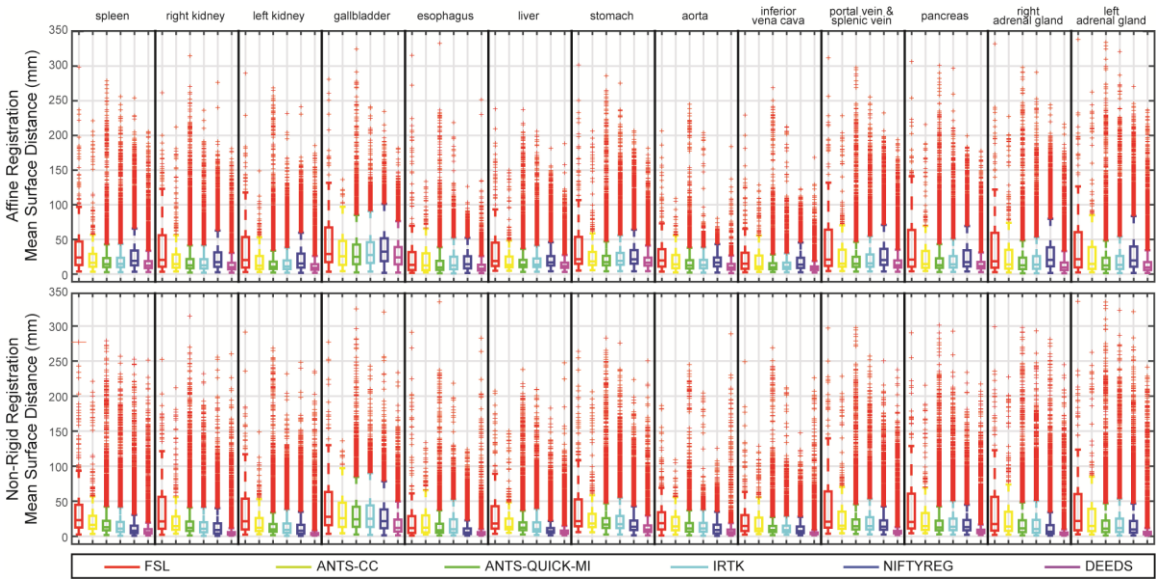

Figure X.4. Boxplot of MSD values on 13 organs for the affine and non-rigid outputs of six registration methods.

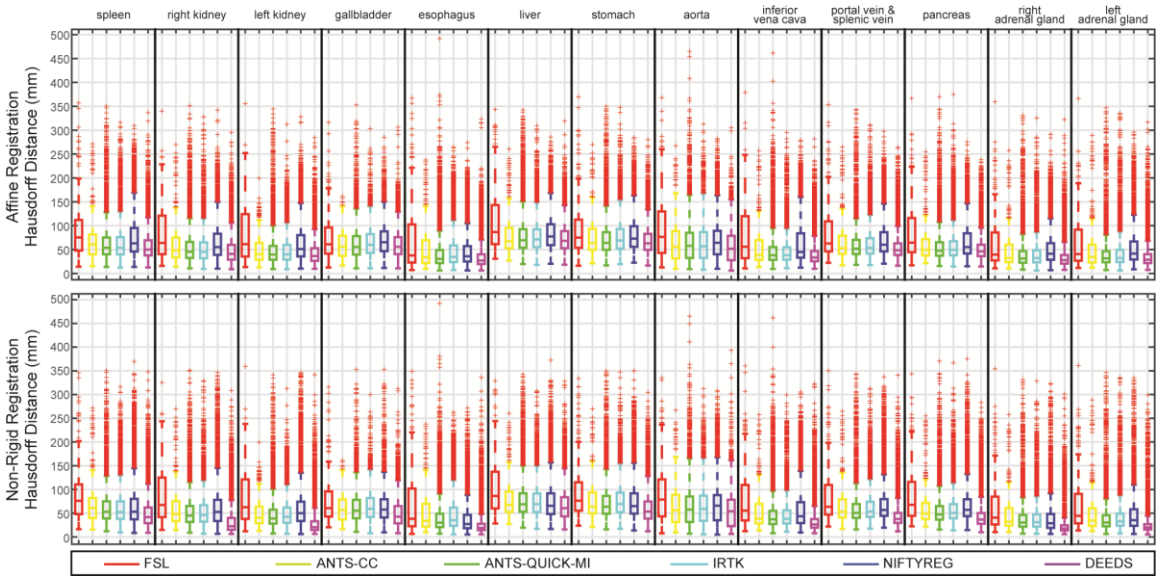

Figure X.5. Boxplot of HD values on 13 organs for the affine and non-rigid outputs of six registration methods.

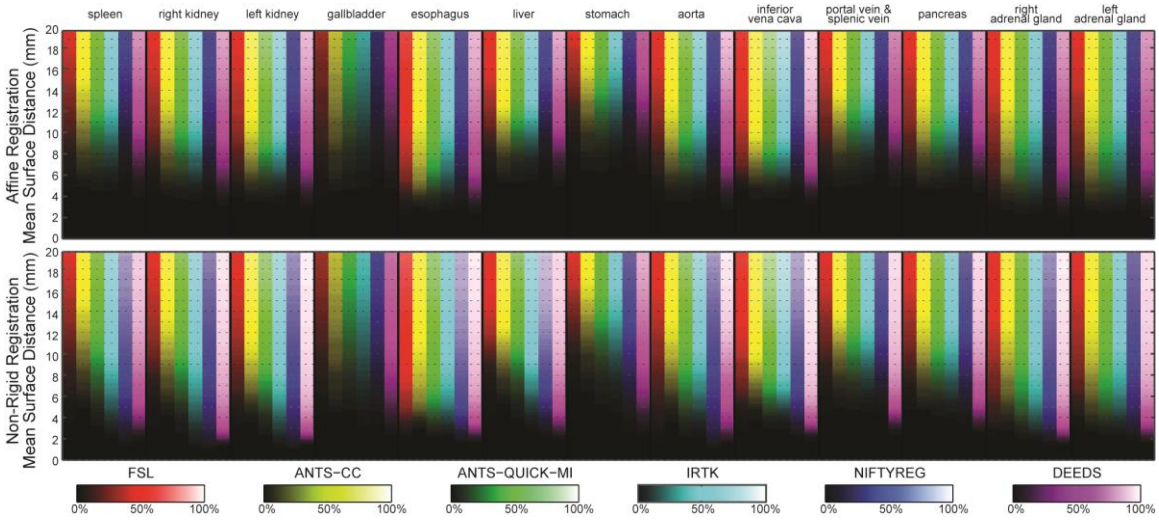

Figure X.6. Brightness-coded cumulative percentages based on MSD values on 13 organs for the affine and non-rigid outputs of six registration methods. Six methods were represented in 6 difference colors. Each column indicates a cumulative curve for the associated organ with the underlying registration method; it demonstrated the percentage of included registration outputs along the increase of the MSD upper bound with its brightness transition from bottom to top. A column with quicker transition from dark to bright indicates more registration outputs with small MSD.

When comparing registration methods with each other, DEEDS presented the best overall DSC of

Table X.1. Metrics on 400 registrations for all tested methods (mean ± std).

Method	DSC	MSD (mm)	HD (mm)	Time (min)
FSL	0.12 ± 0.19	37.92 ± 44.11	84.28 ± 59.96	951.73 ± 201.20
ANTS-CC	0.18 ± 0.21	27.15 ± 32.65	62.92 ± 44.60	411.60 ± 74.20
ANTS-QUICK-MI	0.27 ± 0.25	15.96 ± 19.22	49.66 ± 32.96	50.18 ± 21.93
IRTK	0.28 ± 0.26	19.07 ± 26.50	55.58 ± 39.26	220.27 ± 91.79
NIFTYREG	0.35 ± 0.29	15.72 ± 19.16	59.59 ± 42.60	116.91 ± 34.94
DEEDS	0.49 ± 0.26	8.63 ± 16.16	40.15 ± 32.11	3.73 ± 0.77

Note that ANTS-CC, ANTS-QUICK-MI, and NIFTYREG used two CPU cores for each registration process. The mean DSC across four large organs (liver, spleen, kidneys) is 0.19, 0.31, 0.43, 0.48, 0.55, and 0.70 for FSL, ANTS-CC, ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS, respectively.

Table X.2. Metrics on 9900 registrations for four tested methods (mean ± std).

Method	DSC	MSD (mm)	HD (mm)
ANTS-QUICK-MI	0.23 ± 0.23	20.68 ± 26.14	57.44 ± 39.85
IRTK	0.26 ± 0.26	20.36 ± 24.01	58.71 ± 37.33
NIFTYREG	0.35 ± 0.29	16.98 ± 21.58	62.52 ± 44.29
DEEDS	0.47 ± 0.26	9.79 ± 17.44	43.18 ± 35.08

Note that the mean DSC across 4 large organs (liver, spleen, kidneys) is 0.38, 0.46, 0.55, and 0.68 for ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS, respectively.

affine and non-rigid registration across all organs (**Figure X.3**). For non-rigid registration, NIFTYREG presented slightly higher median DSC over ANTS-QUICK-MI and IRTK, while FSL and ANTS-CC demonstrated overall inferiority comparing to the other three methods. On the MSD and HD boxplots (**Figures X.4, X.5**), the dominance of any registration tool is not visually apparent given the substantial outliers for all methods. To evaluate the results that were not catastrophic failures (i.e., those that could meaningfully contribute to a multi-atlas approach [28, 44]), **Figure X.6** presents MSD results in the form of cumulative percentage, where a higher portion of samples below a certain MSD upper bound was more favorable, where DEEDS yields the most percentage of registrations with lower MSD. **Table X.1** presents the overall performances of DSC, MSD, and HD averaged across all organs for all tested methods on the subset of 400 registrations, while **Table X.2** shows the metrics for ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS on all 9900 registrations; DEEDS demonstrates the best overall performances in both cases. The computation time was also collected in **Table X.1**, where ANTS-QUICK-MI and

Table X.3. Averaged *p*-values of permutation tests between 6 methods performed on 400 registrations.

Method	FSL	ANTS-CC	ANTS-QUICK_MI	IRTK	NIFTYREG	DEEDS
FSL		0.340	0.026	0.057	0.014	0.002
ANTS-CC	0.371		0.098	0.052	0.016	0.001
ANTS-QUICK-MI	0.077	0.266		0.515	0.236	0.010
IRTK	0.216	0.183	0.524		0.249	0.003
NIFTYREG	0.144	0.169	0.517	0.465		0.019
DEEDS	0.032	0.030	0.230	0.044	0.106	

Note the entries in the upper triangular part represent p-values tested on DSC, while those in the lower triangular part were tested on MSD. The shaded entry indicates significant difference ($p < 0.05$) between the correspondent methods of the row and column.

Table X.4. Averaged *p*-values of permutation tests between 4 methods performed on 9900 registrations

Method	ANTS-QUICK-MI	IRTK	NIFTYREG	DEEDS
ANTS-QUICK-MI		0.174	0.000	0.000
IRTK	0.501		0.002	0.000
NIFTYREG	0.255	0.272		0.000
DEEDS	0.024	0.019	0.071	

Note the entries in the upper triangular part represent p-values tested on DSC, while those in the lower triangular part were tested on MSD. The shaded entry indicates significant difference ($p < 0.05$) between the correspondent methods of the row and column.

NIFTYREG could complete in approximately 1h and 2h, respectively using 2 CPU cores, and DEEDS used the least computational time (< 4 min).

The permutation tests found that the superiority of DEEDS in non-rigid registration was significantly better ($p < 0.05$) than all other methods in DSC, and the majority of others in MSD (**Tables X.3, X.4**). The indifference-zone ranking also indicated that DEEDS yielded the best registration performances in an organ-wise manner. NIFTYREG presented the second best results, closely followed by ANTS-QUICK-MI and IRTK, while FSL and ANTS-CC were last (**Figure X.7**).

One registration sample with median overall DSC performances is shown in **Figure X.8**. The volumetric rendering of the registered labels from 6 methods was demonstrated and compared with the manual labels of the target scan to provide a qualitative sense of the registration quality. While large

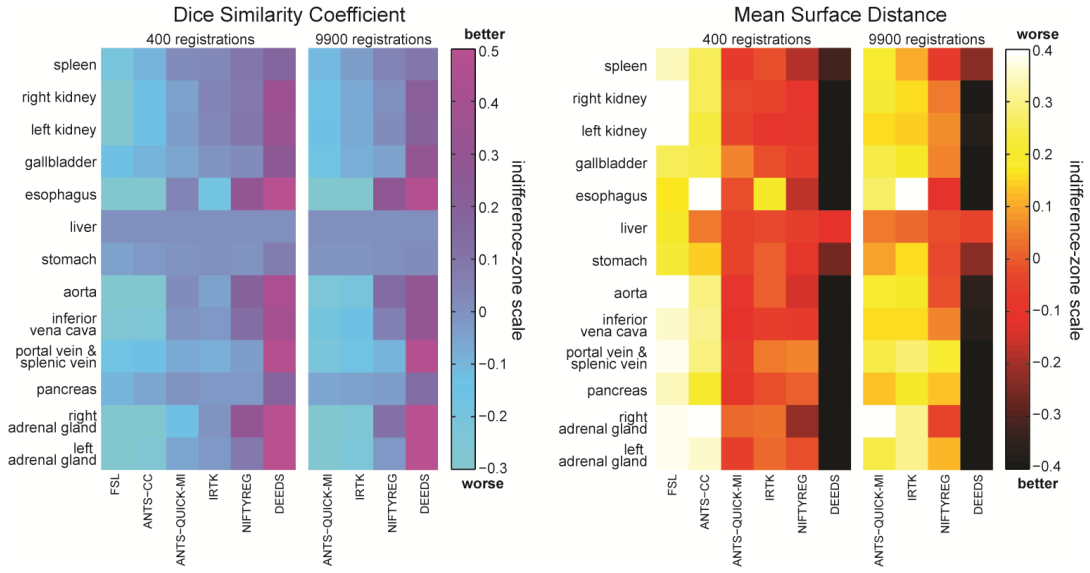

Figure X.7. Indifference-zone map for DSC and MSD. For both metrics, the indifference-zone ranking was applied on 400 registrations for all six methods, and 9900 registrations for ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS. A higher value for the DSC indifference-zone map indicates better performance, while a lower value is more favorable for MSD.

misalignment from all methods can be identified without much effort, ANTS-QUICK-MI, IRTK, NIFTYREG, and DEEDS have the majority of the registered organs located at the close positions, and scaled in the similar sizes with respect to those in the target image. Visually, the organ shapes of target are best captured by DEEDS.

Three pairs of registrations were selected with top 5%, ± 5% around median, and bottom 5 % overall DSC performances, respectively. Registration results on these cases are illustrated in **Figure X.9**, where a coronal slice for each case is selected for the target, source, and all registered images. Based on the overlaid organ labels and the underlying images, DEEDS presents the overall best registrations. Meanwhile, the registration performances are found substantially affected by the similairities between the target and source images including the image FOVs, patient body sizes, organ shapes, and secondary organ complexities (intestines and vessells).

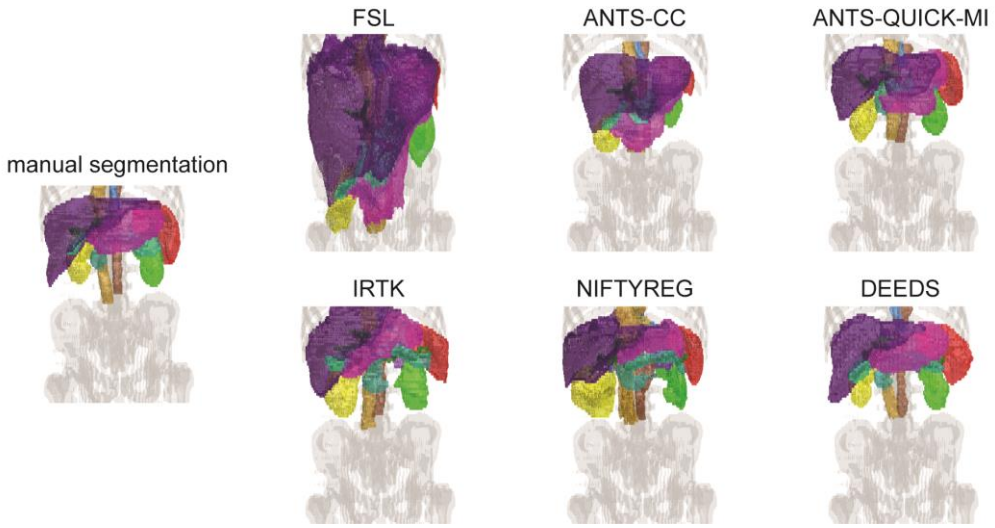

Figure X.8. Volumetric rendering on a single subject with median overall DSC performance. The organ color scheme follows that in Fig. 1.

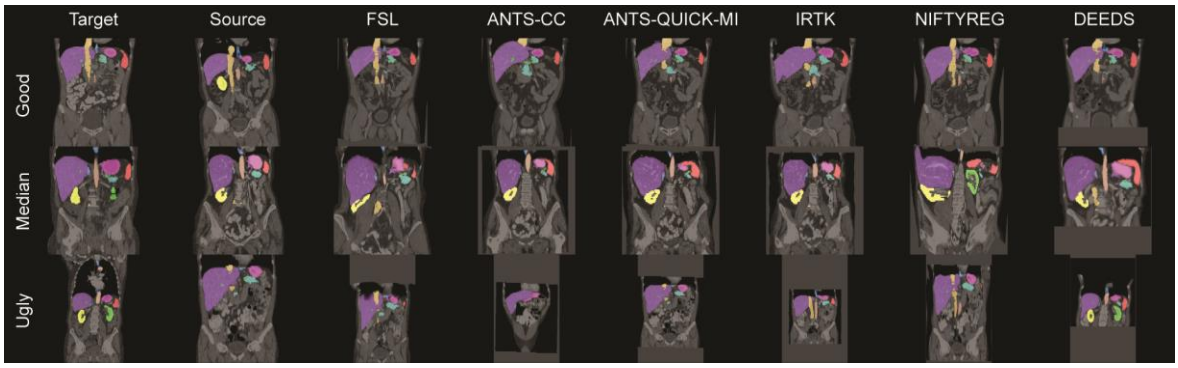

Figure X.9. Illustrations of six registration methods on three registration pairs with good, median, and ugly performances.

4. Discussion

In this study, we analyzed 6 registration methods from 5 different general-purpose image registration toolkits and applied them to abdominal CT scans. Evaluating the volumetric overlap and surface errors on the registered labels on 13 organs of interests showed that the current registration tools were generally far from ideal, where (1) median accuracy was below 0.7 for the majority of organs, and (2) massive outliers indicating catastrophic registration failures were observed. Registration performances are found negatively affected by the dissimilairities between the target and source images including the

image FOVs, patient body sizes, and organ shape, where fundamental body misalignment were observed (**Figure X.9**). Additional challenges come from the implicit discontinuity within the abdomen given the secondary structures (e.g., fat, muscles, bones, intestines in this study). Their variations caused large deformation between different organs of interest so that an affine registration can hardly align all organs at the same time. In addition, their extensive presence and large coverage across the abdomen could mislead the registration algorithms and generate undesirable deformation; for the same reason, small organs could be registered to the secondary structures or other large organs.

We note that the registration results in this study could be biased towards the tested datasets. First, all scans were contrast enhanced, where organs could be more distinguishable from muscle and fat tissue. Registrations between non-contrasted scans may demonstrate additional challenges not shown with our datasets. Second, the population of the patients had more chances to share specific abnormalities, e.g., enlarged spleen and liver, defected abdominal wall. In fact, these patients could also have multiple other diseases, have been treated with different surgical procedures, and demonstrate various other abnormalities (atrophied kidney, missing gallbladder). Thus we consider the registration evaluation on our datasets to be biased towards challenging cases. Datasets among health subjects may yield better registration outcomes. On the other hand, contrasted CT scans on patients with all sorts of abdominal diseases are the most common image format acquired in traditional clinical trials. We thus consider the registration evaluation performed in this study valuable for translational research.

Among the tested registration methods in the presented parameter settings, DEEDS provided the best overall performances, whose median DSC, MSD, and HD for all organs were 0.49, 4.93 mm and 31.72 mm, respectively. The DSC metric is in favor of large structures; small disagreement in small structures can result in large decrease in DSC. Given that, and referred to the literature of abdominal organ segmentation [20, 28, 44], we can consider the ideal DSC value for large (liver, spleen, kidneys), medium (pancreas, stomach, aorta, inferior vena cava), and small (gallbladder, esophagus, portal and splenic vein, adrenal glands) organs to be 0.95, 0.85, and 0.6 respectively. Based on these criteria, even the best registration in this study did not provide good enough outcomes. The massive registration failures

further discouraged the direct individual use of the registration tools in clinical applications. However, if combined with pre-processing and post-processing procedures, registrations with this level of overall accuracy are encouraging and have chances to achieve robust end-point results. Essentially, the multi-atlas techniques can be used to augment local interpretation of abdominal CT scans (e.g., segmentation) by using multiple atlas-to-target registrations. Great care must be taken to account for the registration outliers, where atlas selection [42, 76, 127, 141] and statistical fusion [124, 125, 133] are the keys for robust multi-atlas segmentation (MAS). From the perspective of MAS, registration is the bottleneck, especially in abdomen; a good registration tool can yield better segmentation performances.

Based on the results shown in this study, many opportunities are open for future investigation and development for a registration tool tailored for abdomen.

Firstly, although the presented registration configurations were approved by all the developers of the tested registration methods, further optimization was possible, e.g., in terms of levels of multi-resolution strategy, thresholds of intensity range, use of block matching strategy in affine initialization, regularization on deformation, and etc. Across the tested registrations, a good combination of the similarity metrics (mutual information, cross-correlation, sum of squared distance, and Hamming distances of self-similarity context) and transformation models (B-splines and diffeomorphism) has been covered for deformation, while registrations using other transformation models (e.g., demons [224], optical flow [225]) could be evaluated by experts with these approaches in continuing analysis via the newly released public dataset..

Secondly, contributions in abdominal segmentation also provide some hints in the potential development of abdominal registrations. While using existing registration tools for segmentation, many efforts have been focused on standardizing the abdomen space. Wolz et al. [28] constrained a FOV with 25 cm along the cranial-caudal axis before registration. Linguraru et al. [20] initialized the registration by aligning a single landmark (xiphoid process). Okada et al. [226] and Zhou et al. [227] normalized the abdominal space using pre-segmented diaphragm and rib cage. Recent efforts on organ localizing [187] and organ hierarchical modeling [196] provide the options to minimize the impact of the substantial

registration errors. Piece-wise registrations/segmentations have been demonstrated with better performances than their body-wise counterparts [45, 228]. These pre-processing techniques provide extra features other than intensity-based similarity metrics, and can potentially benefit registrations for capturing the most desirable organ deformation.

Thirdly, we see a new direction in fundamental design for the registration method towards the challenging problems in abdomen. DEEDS yields the best performance in this study, and it is differentiated from other methods mainly by using discrete optimization. Instead of relying on differentiable similarity metric in traditional continuous optimization, DEEDS subdivides the image domain into non-overlapping cubic blocks, and calculates the displacement for each block followed by displacement regularization between blocks. This type of discrete design can capture a large range of potential deformations, and thus coped well with the discontinuous pattern between structures of interest in abdomen. Further exploration in the discrete optimization can be expected to benefit the abdominal registrations.

Lastly, we consider that a structured challenge regarding registration in abdomen using the presented datasets will further boost the development of abdomen-specific and/or registration algorithms. More comprehensive benchmarks to evaluate the efficacy of capturing the abdominal organs will be required to solidify the impact of this potential challenge.

5. Conclusion

This manuscript presents the current state of the registration performances on capturing 13 abdominal organs on CT scans by evaluating six academically popular registration methods without extensive optimizations. By doing this study, we expect to (1) recommend a best registration method to the registration users for their abdomen-related applications, and (2) suggest future directions for registration developers towards more robust and accurate registration algorithms in abdomen. Specifically, DEEDS is currently the best choice for registration users to perform abdominal organ

segmentation. Registration developers can focus on the perspectives of discrete optimization, non-intensity-based feature derivation, and parameter configurations.

CHAPTER XI

IMPROVING MULTI-ATLAS ABDOMINAL ORGAN SEGMENTATION VIA ORGAN-WISE NORMALIZATION

1. Introduction

The major challenge in abdominal organ segmentation on computed tomography (CT) lies in the variations of organ shapes, appearances, and locations among subjects. While recent efforts of multi-atlas segmentation (MAS) achieved reasonable results for abdomen with the concentration on atlas selection [42] and label fusion [28] given substantial registration errors, huge potentials are left to further improve the segmentation via optimizing the target-atlas correspondences.

Abdominal organ segmentation is a problem with intrinsic discontinuities; retrieving the organ entities is of the most interest as opposed to mapping the internal and surrounding tissues. Traditionally, image-based registrations [28, 42] or landmark-based deformation based on pre-segmented structures [18, 27, 196] are commonly applied to align atlas images with a target on a volume basis. Designed to preserve the global topology, these body-wise normalization approaches fail to characterize the intrinsic discontinuities between the organs of interest, and thus result in localization errors that greatly reduce the overall accuracy and robustness.

Organ-wise normalization, on the other hand, handles each organ within its regions of interest (ROI), and could provide more desirable target-atlas correspondences if the organs could be automatically localized (**Figure XI.1**) [187]. Hitherto, the RF organ localization has not been incorporated into a MAS framework.

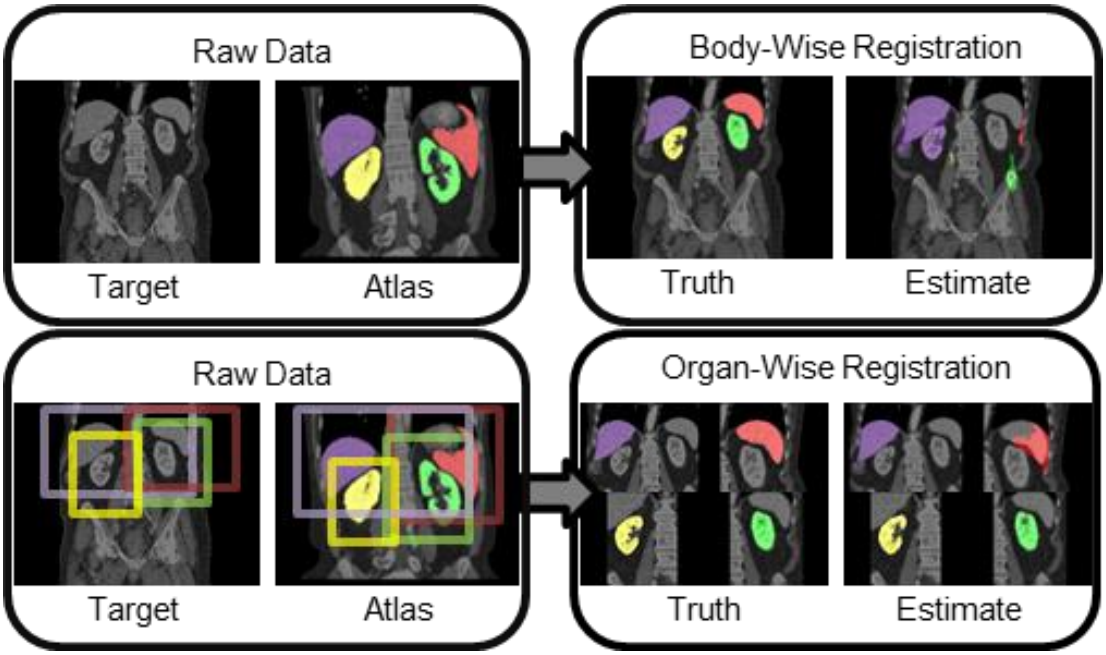

Figure XI.1. (upper lane) Traditional body-wise registration; (lower lane) organ-wise registrations. Note the regions of interest (ROI) on the target are automatically estimated, whereas those on the atlas are derived from the manual segmentation.

The primary contributions of this manuscript are as follows (**Figure XI.2**). (1) We integrate RF organ localization with traditional image-based registration to perform organ-wise normalization. (2) We employ the state-of-the art atlas selection (SIMPLE context learning [42]) and label fusion (joint label fusion, JLF [134]) techniques on the organ-wise normalized atlases. (3) Finally, we combine these contributions to create a new MAS framework to segment 12 abdominal organs on cancer and hernia patients.

2. Methods

2.1. Data

Under Institutional Review Board (IRB) supervision, 30 scans were randomly selected from a combined cohort with metastatic liver cancer and ventral hernias. The 30 scans were captured during portal venous contrast phase with variable volume sizes (512 x 512 x 61 ~ 512 x 512 x 198). The in-plane

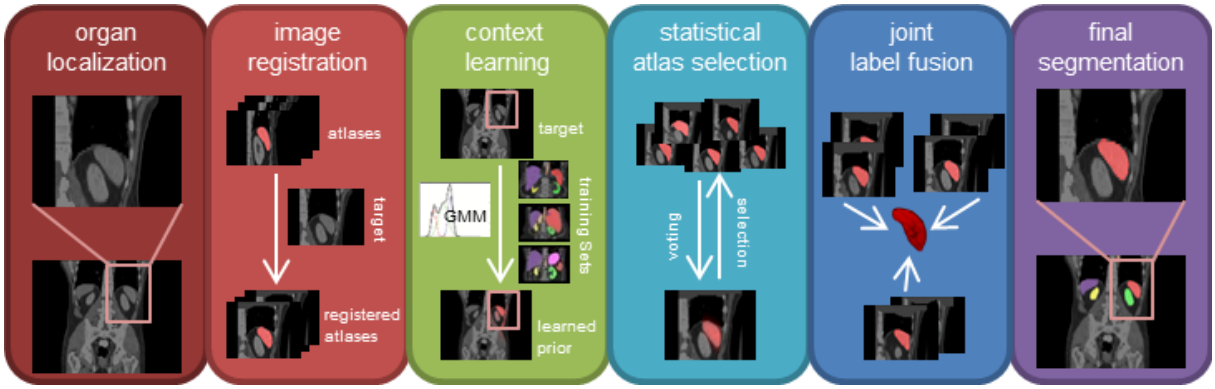

Figure XI.2. Flowchart of the proposed method. Given a target image, organs of interest are first localized using regression forests techniques. Organ-specific cropped atlases are then registered to the estimated ROI using non-rigid image registration. Regularized with a body-wise prior of the target learned through context learning, atlas selection is performed on the registered cropped atlases. Joint label fusion characterizes the correlated errors among the selected atlases, and yields the final segmentation.

resolution varies from 0.59 x 0.59 mm^2 to 0.98 x 0.98 mm^2, while the slice thickness ranges from 2.5 mm to 7.0 mm. Twelve abdominal organs (see in **Figure XI.3**) were manually labeled by two experienced undergraduate students, and verified by a radiologist on a volumetric basis using the MIPAV software (NIH, Bethesda, MD). Ten of the 30 subjects were randomly selected as the training sets for organ localization, and context learning; the remaining 20 were used to validate segmentation results.

2.2. Organ Localization

Organs of interest were localized within bounding boxes following [187]. Briefly, let $\mathcal{M}$ be a set of organ classes, then each organ $m \in \mathcal{M}$ in a CT volume is localized via the boundary positions (in mm) of its bounding box, represented with a six-dimensional vector $\boldsymbol{b}_m = (b_m^L, b_m^R, b_m^A, b_m^P, b_m^H, b_m^F)$. Let $\boldsymbol{u} = (u_x, u_x, u_y, u_y, u_z, u_z)$ indicate the position of each voxel within a CT volume, the voxel displacement towards the boundaries of organ m is then $\boldsymbol{d}_m(\boldsymbol{u}) = \boldsymbol{u} - \boldsymbol{b}_m(\boldsymbol{u})$, where both $\boldsymbol{d}_m$ and $\boldsymbol{b}_m$ are considered as functions of voxels. Consider a regression forest containing T decision trees, where each tree t presents an ensemble of classifiers through its split nodes. Let $f(\boldsymbol{u}; \theta_j) > \tau_j$ be the binary test associated with the

j^{th} node, where $f(\boldsymbol{u};\theta_j)$ and τ_j represent the feature operator and threshold, respectively. Based on the binary tests, the incoming voxels are classified into left and right child nodes recursively, until less than n_{min} voxels are left or a maximum tree depth *D* is reached. Given the randomness of the possible feature operator, a pair of feature operator and threshold to split the sample points is optimized for each node that maximizes the information gain. The leaf nodes are indexed by l across the forest. Here we followed [187] for the choices of feature operator and the criteria of node optimization.

For forest training, given known $\boldsymbol{b}_m$ for all voxels, the training process optimizes the selection of $f(\boldsymbol{u};\theta_j)$ and τ_j for each split node *j* (see section 2.3), and stores the assumed Gaussian distribution of voxel positions $p(\boldsymbol{d}_m|l)$ for each leaf node l in the form of mean $\bar{\boldsymbol{d}}_m$ and covariance Λ_m. In this study, 10 subjects were used for forest training on $T = 4$ trees.

For forest testing, each test voxel $\boldsymbol{u}$ is pushed through all trees with trained binary tests to reach the leaf nodes. Given that $\bar{\boldsymbol{b}}_m(\boldsymbol{u}) = \boldsymbol{u} - \bar{\boldsymbol{d}}_m(\boldsymbol{u})$, the distribution of the absolute boundary position $p(\boldsymbol{b}_m|l)$ can be derived for each organ in leaf l. Typically, a subset of leaf nodes $\tilde{\mathcal{L}}$ with the least uncertainty is used for estimation for each class c. The posterior probability is weighted by the proportion of voxels reaching leaf l, i.e., $p(\boldsymbol{b}_m) = \sum_{l\in\tilde{\mathcal{L}}} p(\boldsymbol{b}_m|l)p(l)$, and the final estimation is given by $\tilde{\boldsymbol{b}}_m = \underset{\boldsymbol{b_m}}{\operatorname{argmax}}\, p(\boldsymbol{b}_m)$.

2.3. Image Registration

Image registrations were applied on organ-specific cropped images using a leave-one-out scheme among the 20 testing subjects. The organ-specific ROIs were defined by padding 5 cm to each side of bounding boxes to include adequate background tissues. The bounding boxes were estimated using RF for the target, while derived from the manual segmentations for the atlases. Affine followed by non-rigid registrations using NiftyReg [188] were performed to align the atlases to the target; both stages involved five coarse-to-fine levels, and limited the upper intensity threshold to 500.

2.4. Context Learning

Different classes of tissues in CT images can be characterized with multi-dimensional Gaussian mixture models using intensity and spatial "context" features. On a voxel-wise basis, let $\boldsymbol{v} \in \mathbb{R}^{d\times 1}$represent a d dimensional feature vector, $m \in \mathcal{M}$ indicate the tissue membership. The probability of the observed features given the tissue is c can be represented with the mixture of N_G Gaussian distributions,

$$f(\boldsymbol{v}|m=t) = \sum_{k=1}^{N_G} \frac{\alpha_{kt}}{(2\pi)^{\frac{d}{2}}|\boldsymbol{C}_{kt}|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(\boldsymbol{v}-\boldsymbol{\mu}_{kt})^T \boldsymbol{C}_{kt}^{-1}(\boldsymbol{v}-\boldsymbol{\mu}_{kt})\right] \quad (11.1)$$

where $\alpha_{kt} \in \mathbb{R}^{1\times 1}$, $\boldsymbol{\mu}_{kt} \in \mathbb{R}^{d\times 1}$, and $\boldsymbol{C}_{kt} \in \mathbb{R}^{d\times d}$ are the unknown mixture probability, mean, and covariance matrix to estimate for each Gaussian mixture component k of each tissue type t by the Expectation-Maximization (EM) algorithm following [107].

The context model can be learned from datasets with known tissue separations, and then the tissue likelihoods on unknown dataset can be inferred by Bayesian expansion and flat tissue membership probability from extracted feature vectors.

$$f(m=t|\boldsymbol{v}) = \frac{f(\boldsymbol{v}|m=t)f(m=t)}{\sum_{t'} f(\boldsymbol{v}|m=t')f(m=t')} = \frac{f(\boldsymbol{v}|m=t)}{\sum_{t'} f(\boldsymbol{v}|m=t')} \quad (11.2)$$

Here, we train context models for 15 tissue types, including 12 manually traced organs, and three automatically retrieved tissues (i.e., muscle, fat, and other) using intensity clustering and excluding the traced organ regions. Six context features were extracted, including intensity, gradient, and local variance, and three spatial coordinates with respect to a single landmark, which was loosely identified as the mid-frontal point of the lung at the plane with the largest cross-sectional lung area. We specified $N_G = 3$. For each organ, the foreground and background likelihoods were learned from the context models based on the context features on target images before cropping, and used as a two-fold spatial prior to regularize the organ-wise atlas selection as follows.

2.5. Statistical Atlas Selection

Organ-specific atlas selection was performed based on the re-formulated SIMPLE [42, 127] algorithm from the perspective of EM. Consider a collection of R registered atlases with label decisions, $\boldsymbol{D} \in \boldsymbol{L}^{N\times R}$, where N is the number of voxels in each registered atlas, and $\boldsymbol{L} = \{0,1,\dots,L-1\}$ represents the label sets. Let $\boldsymbol{c} \in \boldsymbol{S}^{R}$, where $\boldsymbol{S} = \{0,1\}$ indicates the atlas selection decision, i.e., 0 – ignored, and 1 – selected. Let i be the index of voxels, and j of registered atlases. A non-linear rater model, $\theta \in \mathbb{R}^{R\times 2\times L\times L}$, considers the two atlas selection decisions. Let the ignored atlases be no better than random chance, and the selected atlases be slightly inaccurate with error factors $\boldsymbol{\epsilon} \in \mathbf{E}^{\mathrm{R}\times 1}$, where $\mathbf{E} \in (0, \frac{L-1}{L})$. Thus

$$\theta_{j0s's} = \frac{1}{L}, \quad \forall s' \; ; \; \theta_{j1s's} = \begin{cases} 1-\epsilon_j, & s' = s \\ \dfrac{\epsilon_j}{L-1}, & s' \neq s \end{cases} \tag{11.3}$$

where each element $\theta_{jns's}$ represents the probability that the registered atlas j observes label s' given the true label is s and the atlas selection decision is n with an error factor ϵ_j if selected, – i.e., $\theta_{jns's} \equiv f\left(D_{ij} = s' \middle| T_i = s, c_j = n, \epsilon_j\right)$.

Following [125], let $\boldsymbol{W} \in \mathbb{R}^{L\times N}$, where $W_{si}^{(k)}$ represents the probability that the true label associated with voxel i is label s at the k^{th} iteration. Using Bayesian expansion and conditional inter-atlas independence, the E-step can be derived as

$$W_{si}^{(k)} = \frac{f(T_i = s)\prod_j f\left(D_{ij}\middle| T_i = s, c_j^{(k)} = n, \epsilon_j^{(k)}\right)}{\sum_{s'} f(T_i = s')\prod_j f\left(D_{ij}\middle| T_i = s', c_j^{(k)} = n, \epsilon_j^{(k)}\right)} \tag{11.4}$$

where $f(T_i = s)$ is a voxel-wise *a priori* distribution of the underlying segmentation. The exogenous information learned from context features in Eq. 2 can be considered as the Bayesian prior to regularize the atlas selection.

In the M-step, the estimation of the parameters is obtained by maximizing the expected value of the conditional log likelihood function found in Eq. 4. For the error factor,

$$\begin{aligned}\epsilon_j^{(k+1)} &= \arg\max_{\epsilon_j} \sum_i E\left[\ln f\left(D_{ij}\middle|T_i, c_j^{(k)}, \epsilon_j\right)|\boldsymbol{D}, {c_j}^{(k)}, \epsilon_j^{(k)}\right] \\ &= \arg\max_{\epsilon_j} \sum_{s'} \sum_{i:D_{ij}=s'} \sum_s W_{si}^{(k)} \ln \theta_{jc_j^{(k)}s's} \equiv L_{\epsilon_j}\end{aligned} \tag{11.5}$$

Consider the binary segmentation for simplicity, let $M_{TP} = \sum_{i:D_{ij}=1} W_{1i}^{(k)}$,$M_{FP} = \sum_{i:D_{ij}=1} W_{0i}^{(k)}$, $M_{FN} = \sum_{i:D_{ij}=0} W_{1i}^{(k)}$, $M_{TN} = \sum_{i:D_{ij}=0} W_{0i}^{(k)}$, and $\mathrm{M}_T = M_{TP} + M_{TN}$, $M_F = M_{FP} + M_{FN}$. After taking partial derivative of L_{ϵ_j},

$$\epsilon_j^{(k+1)} = \frac{M_F}{M_T + M_F}, i.e., 1 - \epsilon_j^{(k+1)} = \frac{M_T}{M_T + M_F} \tag{11.6}$$

Then for the atlas selection decision

$$\begin{aligned}c_j^{(k+1)} &= \arg\max_{c_j} \sum_i E\left[\ln f\left(D_{ij}\middle|T_i, c_j, \epsilon_j^{(k+1)}\right)|\boldsymbol{D}, {c_j}^{(k)}, \epsilon_j^{(k+1)}\right] \\ &= \arg\max_{c_j} \sum_{s'} \sum_{i:D_{ij}=s'} \sum_s W_{si}^{(k)} \ln \theta_{jc_js's}.\end{aligned} \tag{11.7}$$

Given the intermediate truth estimate $W_{si}^{(k)}$, $c_j^{(k+1)}$ can be maximized by evaluating each 0/1 atlas selection separately. Note the selecting/ignoring behavior in Eq. 7 is parameterized with the error factor ϵ_j, and thus affected by the four summed values of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) as in Eq. 6. An informed prior (e.g., Eq. 2) would greatly deemphasize the TN and yield a metric similar to Dice Similarity Coefficient (DSC) used in the original SIMPLE [127]. We refer this atlas selection technique as SIMPLE context learning, and the majority vote (MV) fusion of its selected atlases as CLSIMPLE.

2.6. Joint Label Fusion

JLF was employed on the selected atlases to yield the final segmentation for each organ. Using the same notation above, JLF defines a voting-based probability for label s as $W_{si} = \sum_{j:D_{ij}=s} w_{ij} D_{ij}$, where w_{ij} indicates the local weight. The local weight $\boldsymbol{w}_i$ is optimized by minimizing the correlated errors between atlases, i.e., $\boldsymbol{w_i}^* = \underset{\boldsymbol{w}_i}{\operatorname{argmin}} \boldsymbol{w}_i{}^t (M_i + \alpha I)\boldsymbol{w}_i$ subject to $\sum_{j=1}^n w_{ij} = 1$ along n atlases, where M_i is an $n \times n$ matrix, estimating the likelihood of correlated errors between each pair of atlases,

and α is a small positive number for regularization. Briefly, M_i can be estimated from the intensity difference on atlas images in a local neighborhood of voxel i, and refined by local patch search. More details can be found in [134].

When using JLF, we specified the local search radii as $3 \times 3 \times 3$, the local patch radii as $2 \times 2 \times 2$, and set the intensity difference mapping parameter (β), and the regularization term (α) as 2 and 0.1, respectively (i.e., default parameters). We refer JLF using the atlases selected by SIMPLE context learning as CLSIMPLEJLF.

2.7. Validation

We quantified the performance of eight MAS approaches in terms of DSC, including MV, JLF, CLSIMPLE, and CLSIMPLEJLF with body-wise and organ-wise normalization. The intermediate results were also examined (not used for segmentations). Specifically, we collected the DSC values for registered atlases, and computed the organ localization errors as the distances between the estimated and true bounding boxes.

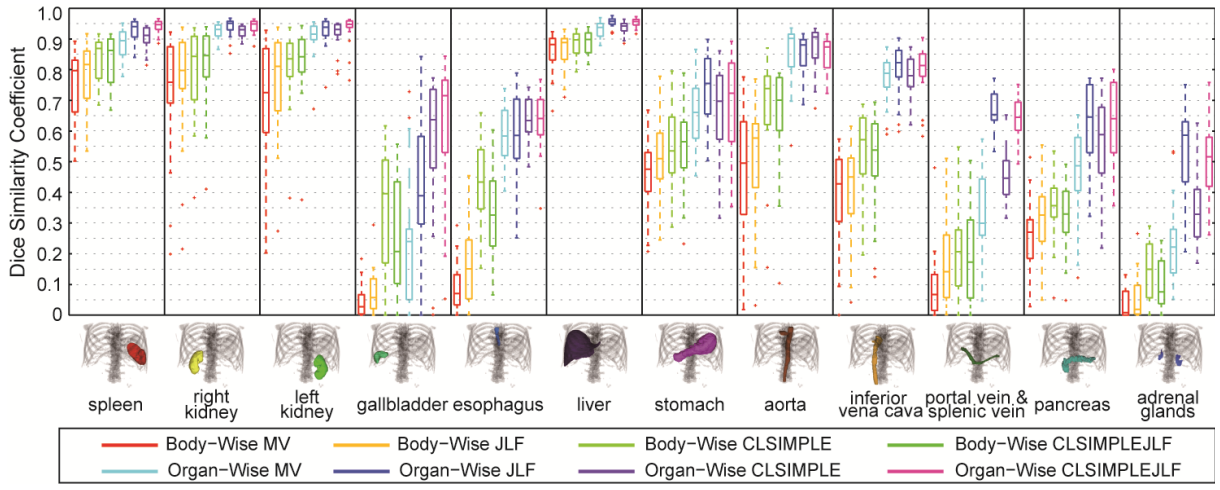

Figure XI.3. Quantitative comparison of in dice similarity coefficient among body-wise and organ-wise label fusion methods. Note the organ color scheme as shown in the rendered tick labels above (e.g., spleen in red, liver in purple) are shared among all figures.

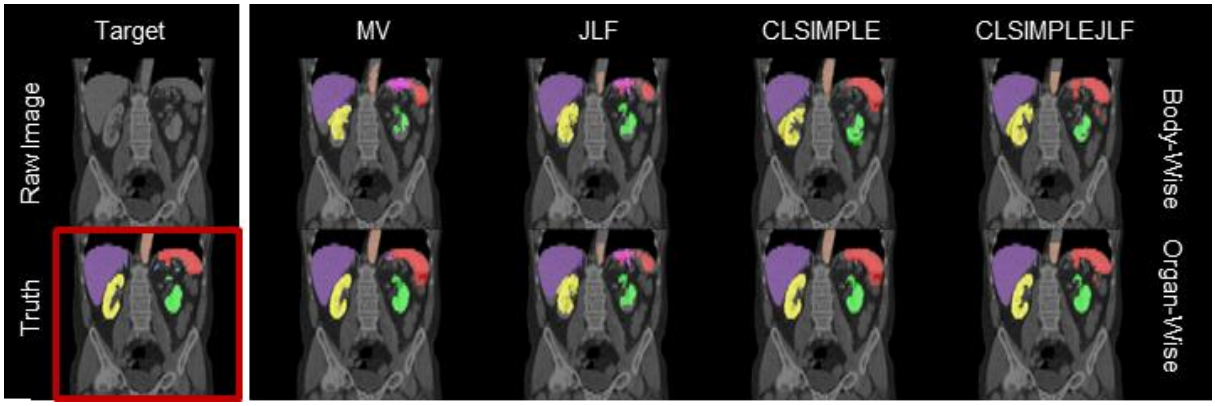

Figure XI.4. Representative qualitative results. Note the subject above was selected as with median overall accuracy in terms of DSC.

3. Results and Discussion

The RF organ localization presented a mean error of 12.83 mm. Body-wise and organ-wise registrations achieved an overall mean DSC of 0.47 and 0.58, respectively. The MAS approaches using organ-wise normalization performed consistently better than their body-wise counterparts over all organs (**Figure XI.3**). The overall mean DSC improvement for CLSIMPLEJLF was 46.2% (0.76 vs. 0.52, $p < 0.001$). Without sophisticated atlas selection and label fusion techniques, organ-wise MV reached an average DSC of 0.91 for spleen, kidneys, and liver. Segmentation failures were substantially removed; MAS methods were benefit from organ-wise normalization for capturing the underlying anatomy more accurately (**Figure XI.4**). In addition, restricting ROI for each organ vastly reduced the computation time (**Table XI.1**) running on a 64-bit 12-core Ubuntu Linux workstation without multi-threading and code optimization. Organ-wise CLSIMPLEJLF presented the best overall DSC among all tested methods ($\Delta = 0.05$ to the second best, $p < 0.001$) with a median DSC over 0.5 for each organ; it also showed competitive DSC values (spleen – 0.94, kidneys – 0.94, liver – 0,95, pancreas – 0.63) comparing to the reported state-of-the-art techniques [18, 27, 28, 196]. In [42], body-wise CLSIMPLEJLF demonstrated better spleen segmentation performances over [28] on the same datasets.

The proposed MAS framework addressed the inter-organ variations with organ-wise normalization in a non-parametric manner; it is generic and applicable to other thoracic, abdominal, and

Table XI.1. Approximated computational time (minutes) for tested label fusion methods.

	MV	JLF	CLSIMPLE	CLSIMPLEJLF
Body-Wise	1.5	360	30	90
Organ-Wise	0.5	240	2	30

Note that registration time is not included.

pelvic organs. This differs from the parametric approaches using statistical shape models [196], where the inter-organ relationships are explicitly characterized along with a pre-defined hierarchy, and a large number of training sets are required. This work creates a bridge between the techniques of high-level identification with low-level extraction for the anatomical structures; refined organ localization and registration-based fusion techniques can further improve the segmentation performances. These results open substantive opportunities for large-scale biomarker screening and image-guided intervention on clinically acquired CT.

CHAPTER XII

CONCLUSION AND FUTURE WORK

1. Summary

The human abdomen is an important compartment that covers the main anatomical structures of multiple biological systems including but not limited to the circulatory (e.g., aorta, inferior vena cava), digestive (e.g., esophagus, stomach, liver, gallbladder, pancreas), urinary (e.g., kidneys), endocrine (adrenal glands), muscular (e.g., rectus and oblique muscles), and lymphatic (e.g., spleen) systems. Computed tomography (CT) scans are acquired in clinical trials as accessible proxies to the human anatomy without opening surgeries, while segmentation of the anatomical structures on CT scans enables quantitative analyses for the physiological and pathological status, and supports the clinical decisions [229]. This dissertation presents automatic segmentation approaches for abdominal wall and abdominal organs on clinically acquired CT. State-of-the-art image processing techniques are investigated and robustly adapted to the challenge problems in abdomen given (1) anatomical structures with substantial occurrences of abnormalities and large variations in shapes and appearances, and (2) CT scans with various image sizes and resolutions, fields of view (FOV), contrast enhancement, and imaging artifacts. Preliminary clinical studies are performed to demonstrate the efficacy of the presented segmentation to assist clinical decisions. Future efforts can build on this work by (1) investigating technical algorithms to improve the segmentation performances, (2) verifying and augmenting the robustness of the segmentation framework on any new datasets for clinical use, (3) translating the structural segmentation into clinically meaningful metrics, and (4) exploring correlation between imaging-based metrics and surgical/radiological outcomes.

2. Abdominal Wall

2.1. Contributions

The abdominal wall is a muscular boundary for the abdominal cavity to protect the internal organs. Its shape, appearance, intactness, and the abdominal compartmentalization are of great clinical interest. The main contributions in terms of abdominal wall in this dissertation include:

1. The first detailed and reproducible anatomical labeling protocol was designed for ventral hernia characterization on CT scans with abdominal wall surfaces, fascia boundaries, and bony landmarks. Shape-, location-, and body-related metrics were derived from the labels and enabled clinical endpoint prediction. This framework also provided objective criteria for clinical ventral hernia analyses, and manual references for the investigation of automatic segmentation (**Chapter III**).
2. An immersive virtual reality system was developed to visualize abdominal CT of ventral hernia patients for better understanding of the abdominal wall and other hernia-related structures with stereoscopic perceptions. Functionalities were designed to allow user adjustable observation of (1) 2-D slices in arbitrary positions/orientations, (2) 3-D volume rendering, and (3) surface rendering of pre-segmented structures (**Chapter IV**).
3. Texture-based features were investigated and integrated into an edge-based level set framework to automatically segment the outer surface of the abdominal wall in the presence of ventral hernia. Improved performances were observed using the texture-based features over the traditional intensity-based features, and achieved an overall mean surface error at 2 mm (**Chapter V**).
4. The first automatic segmentation approach for the whole abdominal wall was developed using an augmented active shape model (AASM) framework that combines multi-atlas label fusion and level set techniques. The whole abdominal wall extracts both the outer and inner surfaces and extends from the xiphoid process to the pubic symphysis; the estimated

segmentation enables the measurement of subcutaneous and visceral fat areas with close correlation to those derived from manual segmentation (**Chapter VI**).

2.2. Future work

Future efforts on the technical and clinical investigation of abdominal wall can be approached from the following perspectives:

1. More sophisticated classification system can be designed to cluster datasets with similar shapes of abdominal walls to improve the AASM segmentation framework. Choices of image features, shape characteristics, and manifold learning algorithms can be investigated to capture the variations of the desirable structures (abdominal wall) as opposed to those of the secondary structures [78, 230].
2. The AASM algorithm can be extended in multiple directions. First, texture-based features can be used for local correspondence search within the AASM framework to improve its robustness of abdominal wall segmentation in the presence of ventral hernias [38]. Second, 3-D version of AASM can be implemented with appropriate treatment of landmark correspondence [111, 193] to benefit the volume-based regularization of the abdominal wall. Third, AASM can also be augmented for multi-structure segmentation, where hierarchical relationship in shapes across structures [181], and multi-channel level set evolution [63] need to be integrated; structures difficult to identify individually can be potentially segmented appropriately together with the abdominal wall (e.g., fascia boundaries).
3. On CT scans of ventral hernia patients, the segmentation of herniated regions remains as an important problem to solve. Possible approaches to extract herniated regions include (1) learning distinguishable features using random forest [231] and deep learning [232] techniques, or (2) referring to literature in detecting other abnormalities like lesions and cancer metastases . The segmentation of the abdominal wall could constrain the searching range for identifying hernias.

4. Metrics of clinical interests need to be robustly derived from the automatic segmentation of the abdominal wall, herniated region, and other associated structures in place of the manual labeling as performed in **Chapter III**. Correlation tests can be performed between the derived metrics and clinically collected records. Given appropriate data setup and experiment design, interesting problems can be studied, e.g., prediction of hernia formation after opening surgeries, estimation of hernia recurrence rate after repair; these potential capabilities would change the treatment procedures drastically for better patient care.

3. Abdominal Organs

3.1. Contributions

The abdominal organs are the key structures to many disease diagnoses and prognoses. Automatic segmentation of the abdominal organs enables shape measurements as quantitative support for large-scale clinical studies. The main contributions in terms of abdominal organs in this dissertation include:

1. A shape constrained framework was presented for automatic spleen segmentation. By combining parametric (active shape model) and non-parametric (multi-atlas label fusion) methods. Spleen shapes models were established based on the signed distance function representation; projected shapes were converted into voxel-wise priors to constrain the multi-atlas segmentation of spleen, and effectively reduce the surface errors (**Chapter VII**).
2. A robust fully automatic framework was developed for segmenting multiple abdominal organs (e.g., spleen, liver, kidneys, and up to 13 organs). A context learned atlas selection approach was proposed to improve the accuracy and efficiency of multi-atlas segmentation given substantial registration errors (**Chapter VIII**).
3. A clinical study was performed on a longitudinal datasets for estimating spleen volumes. A semi-automatic pipeline was proposed for clinical use that deployed automatic computer-assisted segmentation and manual outlier correction. Superior robustness (0.99 in Pearson's

correlation to manual measurement) and efficiency (1 min per scan on average) were demonstrated over other clinical approaches (**Chapter IX**).

4. Six state-of-the-art registration methods were evaluated on abdominal CT scans (used approximately 100,000 hours of CPU time on 100 datasets with cross-validation). Discrete optimization was found effective to capture the implicit discontinuities between structures of interest in abdomen, and thus suggested for the investigation of registration tools tailored for abdomen (**Chapter X**).
5. Random forest techniques were investigated on localizing abdominal organs on CT scans. Organ-wise multi-atlas labeling was then performed on the localized regions of interest. Using organ-wise labeling, consistent improvement was observed over the traditional multi-atlas framework, less computational time was also achieved (**Chapter XI**).

3.2. Future work

Future investigation on abdominal organs can be approached from the following technical and clinical perspectives:

1. The pipeline of multi-atlas abdominal organ segmentation presented in **Chapters VIII** and **XI** can achieve registration methods better tailored for abdomen. While easily achievable improvement can be expected by using the best available registration tool presented in **Chapter X**, further development of abdomen-specific registration towards discrete optimization can boost the abdominal segmentation fundamentally. Meanwhile, GPU implementation of the registration and label fusion techniques [188, 233] should be considered to reduce the computational cost of multi-atlas segmentation that potentially enables near real-time automatic segmentation.
2. The random forest localization for abdominal organs is useful, but far from ideal given that (1) it is sensitive to datasets with various FOVs, (2) it only provides a loose constraint with its bounding box representation. Robust segmentation of related bony structures (e.g., rib cages,

spinal vertebrae, pelvic bones) can be investigated to provide consistent references for organ localization regardless of FOVs, and anatomically sensible organ boundaries when combined with atlas-based registrations [46, 234, 235].

3. Segmentation refinement, especially on small organs (e.g., gallbladder, adrenal glands) needs to be investigated. Potential directions include but are not limited to (1) corrective learning [236, 237], (2) multi-modal, multi-contrast enhancement [26, 27], (3) label correction using random forest, convolutional neural networks on the basis of voxels, patches or supervoxels [231, 238], and (4) details adjustment using level set, active shape models, and graph cuts [24, 25, 239].
4. The robustness of the volumetric estimation from segmentations needs to be validated. With minimal human intervention in quality assurance, a progressive feedback loop can be used for augmenting the atlas pool and thus the segmentation performances on new datasets. Correlation between the volumetric estimation and clinically recorded metrics can be explored to study pathological inferences and drug deliver effects.

4. Concluding Remarks

In the past five years, an increasing number of researches have begun to explore human abdomen segmentation through development of tailored image processing, computer vision, and machine learning techniques. Many of those techniques originally designed for brains are now being adapted to abdominal structures, where challenges coming from the larger variations within abdomen must be addressed. This dissertation presents automatic segmentation of the major abdominal structures on clinically acquired CT scans with the state-of-the-art performances, provides technical, clinical, and educational impact on understanding the abdominal anatomy, and sets a foundation for future development towards computer-assisted clinical decision support.

APPENDIX A

PUBLICATIONS

1. Journal Articles

1) **Zhoubing Xu**, Andrew J. Asman, Rebeccah B. Baucom, Richard G. Abramson, Benjamin K. Poulose, Bennett A. Landman. "*Quantitative CT Imaging of Ventral Hernias: Preliminary Validation of an Anatomical Labeling Protocol*", PLOS One. 10(10), e0141671 (2015)
2) **Zhoubing Xu**, Ryan P. Burke, Christopher P. Lee, Rebeccah B. Baucom, Benjamin K. Poulose, Richard G. Abramson, Bennett A. Landman "*Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning*", Medical Image Analysis, 24(1), pp. 18-27, (2015).
3) Frederick W. Bryan, **Zhoubing Xu**, Andrew J. Asman, Wade M. Allen, Daniel S. Reich, and Bennett A. Landman. "*Self-Assessed Performance Improves Statistical Fusion of Image Labels*", Medical Physics. 40, 031903 (2014)
4) **Zhoubing Xu**, Wade M. Allen, Rebeccah B. Baucom, Benjamin K. Poulose, Bennett A. Landman, "*Texture Analysis Improves Level Set Segmentation of the Anterior Abdominal Wall*", Medical Physics. 40, 121901 (2013)
5) **Zhoubing Xu**, Andrew J. Asman, Eesha Singh, Lola Chambless, Reid Thompson, and Bennett A. Landman. "*Segmentation of Malignant Gliomas through Remote Collaboration and Statistical Fusion*", Medical physics 39 (2012): 5981.

2. Highly Selective Conference Publications

1) **Zhoubing Xu**, Andrew J. Asman, Peter L. Shanahan, Richard G. Abramson, Bennett A. Landman. *"SIMPLE Is a Good Idea (and Better with Context Learning)",* Conditionally accepted to

International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Boston, USA, September 2014

3. Conference Publications

1) **Zhoubing Xu**, Adam L. Gertz, Rebeccah B. Baucom, Benjamin K. Poulose, Richard G. Abramson, Bennett A. Landman. "Improving Multi-Atlas Abdominal Organ Segmentation via Organ-Wise Normalization." In MICCAI 2015 Workshop and Challenge: Multi-Atlas Labeling Beyond Cranial Vault. Munich, Germany, October 2015
2) **Zhoubing Xu**, Ryan P. Burke, Christopher P. Lee, Rebeccah B. Baucom, Benjamin K. Poulose, Richard G. Abramson, Bennett A. Landman. "Efficient Abdominal Segmentation on Clinically Acquired CT with SIMPLE Context Learning." In Proceedings of the SPIE Medical Imaging Conference. Orlando, Florida, February 2015.
3) Christopher P. Lee, **Zhoubing Xu**, Ryan P. Burke, Rebeccah B. Baucom, Benjamin K. Poulose, Richard G. Abramson, Bennett A. Landman. "Evaluation of Five Image Registration Tools: Pitfalls and Opportunities with Soft Anatomy." In Proceedings of the SPIE Medical Imaging Conference. Orlando, Florida, February 2015.
4) Ryan P. Burke, **Zhoubing Xu**, Christopher P. Lee, Rebeccah B. Baucom, Benjamin K. Poulose, Richard G. Abramson, Bennett A. Landman. "Multi-Atlas Segmentation for Abdominal Organs with Gaussian Mixture Models." In Proceedings of the SPIE Medical Imaging Conference. Orlando, Florida, February 2015.
5) **Zhoubing. Xu**, Bo Li, Swetasudha. Panda, Andrew J. Asman, Kristen. L. Merkle, Peter L. Shanahan, Richard G. Abramson, and Bennett A. Landman. "*Shape-Constrained Multi-Atlas Segmentation of Spleen in CT*". In Proceedings of the SPIE Medical Imaging Conference. San Diego, California, February 2014.
6) Bo Li, Swetasudha Panda, **Zhoubing Xu**, Andrew J. Asman, Peter L. Shanahan, Richard G. Abramson, Bennett A. Landman. "*Regression forest region recognition enhances multi-atlas spleen*

labeling". In Proceedings of the Medical Image Computing and Computer Assisted Intervention (MICCAI) Challenge Workshop on Segmentation: Algorithms, Theory and Applications. Nagoya, Japan, September 2013

7) **Zhoubing Xu**, Wade M. Allen, Benjamin K. Poulose, and Bennett A. Landman. "*Automatic Segmentation of Abdominal Wall in Ventral Hernia CT: A Pilot Study*". In Proceedings of SPIE Medical Imaging Conference, Orlando, Florida, February 2013, pp. 86693T-86693T.
8) Wade M. Allen, **Zhoubing Xu**, Andrew J. Asman, Benjamin K. Poulose, Bennett A. Landman. "*Quantitative Anatomical Labeling of the Anterior Abdominal Wall*". In Proceedings of the SPIE Medical Imaging Conference, Orlando, Florida, February 2013, pp. 867312-867312.
9) Qiufeng Lin, **Zhoubing Xu**, Bo Li, Rebeccah Baucom, Benjamin Poulose, Bennett A. Landman, and Robert E. Bodenheimer. "*Immersive Virtual Reality for Visualization of Abdominal CT*". In Proceedings of the SPIE Medical Imaging Conference, Orlando, Florida, February 2013, pp. 867317-867317.
10) **Zhoubing Xu**, Andrew J. Asman, Eesha Singh, Lola Chambless, Reid Thompson, and Bennett A. Landman, "*Collaborative Labeling of Malignant Glioma*", In Proceedings of the 2012 International Symposium on Biomedical Imaging (ISBI). Barcelona, Spain, May 2012, pp. 1148-1151.
11) **Zhoubing Xu**, Andrew J. Asman and Bennett A. Landman. "*Generalized Statistical Label Fusion using Multiple Consensus Levels*". In Proceedings of the SPIE Medical Imaging Conference. San Diego, California, February 2012, pp. 831411-831411.
12) Eesha Singh, Andew J. Asman, **Zhoubing Xu**, Lola Chambless, Reid Thompson and Bennett A. Landman. "*Collaborative Labeling of Malignant Glioma with WebMILL: A First Look*". In Proceedings of the SPIE Medical Imaging Conference. San Diego, California, February 2012, pp. 831813-1.

4. Abstracts

1) Rebeccah B. Baucom, **Zhoubing Xu**, Wade M. Allen, Andrew J. Asman, Bennett A. Landman, Benjamin K. Poulose. "*The Quantitative Evaluation of Abdominal Wall Hernias Using Anatomical Labeling*". Americas Hernia Society (AHS). Las Vegas, Nevada, March 2014.

REFERENCES

1. Muysoms, F., et al., *Classification of primary and incisional abdominal wall hernias.* Hernia, 2009. **13**(4): p. 407-414.
2. Yao, S., et al., *Significance of measurements of herniary area and volume and abdominal cavity volume in the treatment of incisional hernia: Application of CT 3D reconstruction in 17 cases.* Computer Aided Surgery, 2012. **17**(1): p. 40-45.
3. Sabbagh, C., et al., *Peritoneal volume is predictive of tension-free fascia closure of large incisional hernias with loss of domain: a prospective study.* Hernia, 2011. **15**(5): p. 559-565.
4. DuBay, D.A., et al., *Incisional herniation induces decreased abdominal wall compliance via oblique muscle atrophy and fibrosis.* Annals of surgery, 2007. **245**(1): p. 140.
5. Luijendijk, R.W., et al., *A comparison of suture repair with mesh repair for incisional hernia.* The New England journal of medicine, 2000. **343**(6): p. 392-8.
6. Mazzara, G.P., et al., *Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation.* International Journal of Radiation Oncology* Biology* Physics, 2004. **59**(1): p. 300-312.
7. Landman, B.A., et al., *Foibles, follies, and fusion: web-based collaboration for medical image labeling.* NeuroImage, 2012. **59**(1): p. 530-9.
8. Xu, Z., et al., *Segmentation of malignant gliomas through remote collaboration and statistical fusion.* Medical physics, 2012. **39**(10): p. 5981-9.
9. Bryan, F.W., et al., *Self-assessed performance improves statistical fusion of image labels.* Medical physics, 2014. **41**(3): p. 031903.
10. Fischl, B., et al., *Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain.* Neuron, 2002. **33**(3): p. 341-55.

11. Collins, D.L., et al., *Automatic 3-D model-based neuroanatomical segmentation.* Human Brain Mapping, 1995. **3**(3): p. 190-208.

12. Fischl, B., et al., *Automatically parcellating the human cerebral cortex.* Cerebral Cortex, 2004. **14**(1): p. 11-22.

13. Heckemann, R.A., et al., *Automatic anatomical brain MRI segmentation combining label propagation and decision fusion.* NeuroImage, 2006. **33**(1): p. 115-126.

14. Dickerson, B.C. and R.A. Sperling, *Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease.* NeuroRx, 2005. **2**(2): p. 348-360.

15. Bielekova, B. and R. Martin, *Development of biomarkers in multiple sclerosis.* Brain : a journal of neurology, 2004. **127**(7): p. 1463-1478.

16. Park, H., P.H. Bland, and C.R. Meyer, *Construction of an abdominal Probabilistic atlas and its application in segmentation.* IEEE transactions on medical imaging, 2003. **22**(4): p. 483-492.

17. Zhou, X., et al. *Construction of a probabilistic atlas for automated liver segmentation in non-contrast torso CT images*. in *International Congress Series*. 2005. Elsevier.

18. Shimizu, A., et al., *Segmentation of multiple organs in non-contrast 3D abdominal CT images.* International journal of computer assisted radiology and surgery, 2007. **2**(3-4): p. 135-142.

19. Shimizu, A., et al., *Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography.* International journal of computer assisted radiology and surgery, 2010. **5**(1): p. 85-98.

20. Linguraru, M.G., et al., *Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation.* Medical physics, 2010. **37**(2): p. 771-783.

21. Soler, L., et al., *Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery.* Computer aided surgery : official journal of the International Society for Computer Aided Surgery, 2001. **6**(3): p. 131-42.

22. Heimann, T., I. Wolf, and H.-P. Meinzer, *Active shape models for a fully automated 3D segmentation of the liver–an evaluation on clinical data*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*2006, Springer. p. 41-48.

23. Okada, T., et al., *Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*2008, Springer. p. 502-509.

24. Bagci, U., X. Chen, and J.K. Udupa, *Hierarchical scale-based multiobject recognition of 3-D anatomical structures.* Medical Imaging, IEEE Transactions on, 2012. **31**(3): p. 777-789.

25. Chen, X., et al., *Medical image segmentation by combining graph cuts and oriented active appearance models.* Image Processing, IEEE Transactions on, 2012. **21**(4): p. 2035-2046.

26. Linguraru, M.G., et al., *Multi-organ segmentation from multi-phase abdominal CT via 4D graphs using enhancement, shape and location optimization*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*2010, Springer. p. 89-96.

27. Linguraru, M.G., et al., *Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT.* Medical image analysis, 2012. **16**(4): p. 904-914.

28. Wolz, R., et al., *Automated abdominal multi-organ segmentation with subject-specific atlas generation.* IEEE transactions on medical imaging, 2013. **32**(9): p. 1723-30.

29. Zhang, W., et al. *Segmenting the thoracic, abdominal and pelvic musculature on CT scans combining atlas-based model and active contour model*. in *SPIE Medical Imaging*. 2013. International Society for Optics and Photonics.

30. Chung, H., et al. *Automated segmentation of muscle and adipose tissue on CT images for human body composition analysis*. in *SPIE Medical Imaging*. 2009. International Society for Optics and Photonics.

31. Ding, F., W.K. Leow, and S. Venkatesh. *Removal of abdominal wall for 3D visualization and segmentation of organs in CT volume*. in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. 2009. IEEE.

32. Yao, J., D.L. Sussman, and R.M. Summers, *Fully automated adipose tissue measurement on abdominal CT*, in *Proceedings of SPIE Medical Imaging*2011, International Society for Optics and Photonics. p. 79651Z-79651Z-6.

33. Zhu, W., et al., *Fast segmentation of abdominal wall: application to sliding effect removal for non-rigid registration*, in *Proceedings of the 4th international conference on Abdominal Imaging: computational and clinical applications*2012, Springer-Verlag: Nice, France. p. 198-207.

34. Allen, W.M., et al. *Quantitative Anatomical Labeling of the Anterior Abdominal Wall*. in *Proceedings of SPIE Medical Imaging*. 2013. Orlando, Florida.

35. Xu, Z., et al., *Quantitative CT Imaging of Ventral Hernias: Preliminary Validation of an Anatomical Labeling Protocol.* PloS one, 2015. **10**(10): p. e0141671.

36. Lin, Q., et al., *Immersive Virtual Reality for Visualization of Abdominal CT.* Proceedings of SPIE, 2013. **8673**.

37. Xu, Z., et al., *Automatic Segmentation of Abdominal Wall in Ventral Hernia CT: A Pilot Study.* Proceedings of SPIE, 2013. **8669**.

38. Xu, Z., et al., *Texture analysis improves level set segmentation of the anterior abdominal wall.* Medical physics, 2013. **40**(12): p. 121901.

39. Xu, Z., et al. *Shape-constrained multi-atlas segmentation of spleen in CT*. in *SPIE Medical Imaging*. 2014. International Society for Optics and Photonics.

40. Burke, R.P., et al. *Multi-atlas segmentation for abdominal organs with Gaussian mixture models*. in *SPIE Medical Imaging*. 2015. International Society for Optics and Photonics.

41. Lee, C.P., et al. *Evaluation of five image registration tools for abdominal CT*. in *SPIE Medical Imaging*. 2015. International Society for Optics and Photonics.

42. Xu, Z., et al., *SIMPLE IS a Good Idea (and Better with Context Learning)*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*2014, Springer. p. 364-371.

43. Xu, Z., et al. *Efficient abdominal segmentation on clinically acquired CT with SIMPLE context learning*. in *SPIE Medical Imaging*. 2015. International Society for Optics and Photonics.

44. Xu, Z., et al., *Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning.* Medical image analysis, 2015.

45. Xu, Z., et al. *Improving Multi-Atlas Abdominal Organ Segmentation via Organ-Wise Normalization*. in *MICCAI 2015 Workshop and Challenge: Multi-Atlas Labeling Beyond Cranial Vault* 2015.

46. Yao, J., S.D. O'Connor, and R.M. Summers. *Automated spinal column extraction and partitioning*. in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*. 2006. IEEE.

47. Yao, J. and R.M. Summers, *Statistical location model for abdominal organ localization*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*2009, Springer. p. 9-17.

48. Xu, C. and J.L. Prince, *Snakes, shapes, and gradient vector flow*. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 1998. **7**(3): p. 359-69.

49. Kass, M., A. Witkin, and D. Terzopoulos, *Snakes: Active contour models.* International Journal of Computer Vision, 1988. **1**(4): p. 321-331.

50. Caselles, V., et al., *A Geometric Model for Active Contours in Image-Processing.* Numerische Mathematik, 1993. **66**(1): p. 1-31.

51. Caselles, V., R. Kimmel, and G. Sapiro, *Geodesic active contours.* International Journal of Computer Vision, 1997. **22**(1): p. 61-79.

52. Kichenassamy, S., et al. *Gradient flows and geometric active contour models*. in *Proceedings of the 5th International Conference on Computer Vision*. 1995. IEEE.

53. Tuceryan, M. and A.K. Jain, *Texture analysis.* Handbook of pattern recognition and computer vision, 1993. **276**.

54. Castellano, G., et al., *Texture analysis of medical images.* Clinical radiology, 2004. **59**(12): p. 1061-9.

55. Herlidou, S., et al., *Comparison of automated and visual texture analysis in MRI: characterization of normal and diseased skeletal muscle.* Magnetic resonance imaging, 1999. **17**(9): p. 1393-7.

56. Elfadel, I.M. and R.W. Picard, *Gibbs random fields, cooccurrences, and texture modeling.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1994. **16**(1): p. 24-37.

57. Haralick, R.M., Shanmuga.K, and I. Dinstein, *Textural Features for Image Classification.* Ieee Transactions on Systems Man and Cybernetics, 1973. **Smc3**(6): p. 610-621.

58. Daugman, J.G., *Two-dimensional spectral analysis of cortical receptive field profiles.* Vision research, 1980. **20**(10): p. 847-856.

59. Turner, M.R., *Texture discrimination by Gabor functions.* Biological Cybernetics, 1986. **55**(2-3): p. 71-82.

60. Jain, A.K. and F. Farrokhnia, *Unsupervised texture segmentation using Gabor filters.* Pattern recognition, 1991. **24**(12): p. 1167-1186.

61. Bovik, A.C., M. Clark, and W.S. Geisler, *Multichannel Texture Analysis Using Localized Spatial Filters.* IEEE transactions on pattern analysis and machine intelligence, 1990. **12**(1): p. 55-73.

62. Dunn, D. and W.E. Higgins, *Optimal Gabor filters for texture segmentation.* Image Processing, IEEE Transactions on, 1995. **4**(7): p. 947-964.

63. Brox, T. and J. Weickert, *Level set based image segmentation with multiple regions.* Pattern recognition, 2004. **3175**: p. 415-423.

64. Rousson, M., T. Brox, and R. Deriche. *Active unsupervised texture segmentation on a diffusion based feature space*. in *Proceeding of IEEE computer society conference on Computer vision and pattern recognition*. 2003. IEEE.

65. Karoui, I., et al. *Region-based image segmentation using texture statistics and level-set methods*. in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. 2006. IEEE.

66. Cootes, T.F., et al., *Active Shape Models - Their Training and Application.* Computer vision and image understanding, 1995. **61**(1): p. 38-59.

67. Jolliffe, I., *Principal component analysis*2005: Wiley Online Library.

68. Cootes, T.F., G.J. Edwards, and C.J. Taylor, *Active appearance models.* IEEE transactions on pattern analysis and machine intelligence, 2001. **23**(6): p. 681-685.

69. Wolz, R., et al. *Segmentation of subcortical structures and the hippocampus in brain MRI using graph-cuts and subject-specific a-priori information*. in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. 2009. IEEE.

70. Criminisi, A., et al., *Regression forests for efficient anatomy detection and localization in CT studies*, in *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*2011, Springer. p. 106-117.

71. Zhou, X., et al., *Automatic localization of solid organs on 3D CT images by a collaborative majority voting decision based on ensemble learning.* Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society, 2012. **36**(4): p. 304-13.

72. Perona, P. and J. Malik, *Scale-space and edge detection using anisotropic diffusion.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1990. **12**(7): p. 629-639.

73. Descoteaux, M., et al., *Bone enhancement filtering: application to sinus bone segmentation and simulation of pituitary surgery*. Computer Aided Surgery, 2006. **11**(5): p. 247-255.

74. Hill, D.L., et al., *Medical image registration.* Physics in medicine and biology, 2001. **46**(3): p. R1.

75. Maintz, J. and M.A. Viergever, *A survey of medical image registration.* Medical image analysis, 1998. **2**(1): p. 1-36.

76. Aljabar, P., et al., *Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy*. Neuroimage, 2009. **46**(3): p. 726-738.

77. Artaechevarria, X., A. Munoz-Barrutia, and C. Ortiz-de-Solorzano, *Combination strategies in multi-atlas image segmentation: Application to brain MR data.* Medical Imaging, IEEE Transactions on, 2009. **28**(8): p. 1266-1277.

78. Wolz, R., et al., *LEAP: learning embeddings for atlas propagation.* Neuroimage, 2010. **49**(2): p. 1316-1325.

79. Fitzpatrick, J.M., J.B. West, and C.R. Maurer Jr, *Predicting error in rigid-body point-based registration.* Medical Imaging, IEEE Transactions on, 1998. **17**(5): p. 694-702.

80. Lewis, J. *Fast normalized cross-correlation*. in *Vision interface*. 1995.

81. Cover, T.M. and J.A. Thomas, *Elements of information theory*2012: John Wiley & Sons.

82. Vajda, I., *Theory of statistical inference and information*1989: Kluwer Academic Publishers Dordrecht.

83. Pennec, X., P. Cachier, and N. Ayache. *Understanding the "demon's algorithm": 3D non-rigid registration by gradient descent*. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI'99*. 1999. Springer.

84. Powell, M.J., *An efficient method for finding the minimum of a function of several variables without calculating derivatives.* The computer journal, 1964. **7**(2): p. 155-162.

85. Lagarias, J.C., et al., *Convergence properties of the Nelder--Mead simplex method in low dimensions.* SIAM Journal on Optimization, 1998. **9**(1): p. 112-147.

86. Meijering, E., *A chronology of interpolation: From ancient astronomy to modern signal and image processing.* Proceedings of the Ieee, 2002. **90**(3): p. 319-342.

87. Wahba, G., *Spline interpolation and smoothing on the sphere.* SIAM Journal on Scientific and Statistical Computing, 1981. **2**(1): p. 5-16.

88. Besl, P.J. and N.D. McKay. *Method for registration of 3-D shapes*. in *Robotics-DL tentative*. 1992. International Society for Optics and Photonics.

89. Arun, K.S., T.S. Huang, and S.D. Blostein, *Least-squares fitting of two 3-D point sets.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1987(5): p. 698-700.

90. Barnea, D.I. and H.F. Silverman, *A class of algorithms for fast digital image registration.* Computers, IEEE Transactions on, 1972. **100**(2): p. 179-186.

91. Maes, F., et al., *Multimodality image registration by maximization of mutual information.* Medical Imaging, IEEE Transactions on, 1997. **16**(2): p. 187-198.

92. Nomizu, K., *Affine differential geometry: geometry of affine immersions*1994: Cambridge University Press.

93. Klein, A., et al., *Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.* Neuroimage, 2009. **46**(3): p. 786-802.

94. Bookstein, F.L., *Principal Warps - Thin-Plate Splines and the Decomposition of Deformations.* IEEE transactions on pattern analysis and machine intelligence, 1989. **11**(6): p. 567-585.

95. Rueckert, D., et al., *Nonrigid registration using free-form deformations: application to breast MR images.* Medical Imaging, IEEE Transactions on, 1999. **18**(8): p. 712-721.

96. Avants, B.B., et al., *Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain.* Medical image analysis, 2008. **12**(1): p. 26-41.

97. Otsu, N., *A threshold selection method from gray-level histograms.* Automatica, 1975. **11**(285-296): p. 23-27.

98. Sahoo, P.K., S. Soltani, and A. Wong, *A survey of thresholding techniques.* Computer vision, graphics, and image processing, 1988. **41**(2): p. 233-260.

99. Vincent, L. and P. Soille, *Watersheds in digital spaces: an efficient algorithm based on immersion simulations.* IEEE transactions on pattern analysis and machine intelligence, 1991. **13**(6): p. 583-598.

100. 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, 2004. **26**(9): p. 1124-1137.

101. Boykov, Y., O. Veksler, and R. Zabih, *Fast approximate energy minimization via graph cuts.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001. **23**(11): p. 1222-1239.

102. Bishop, C.M., *Pattern recognition and machine learning*. Vol. 1. 2006: springer New York.

103. Zhang, Y., M. Brady, and S. Smith, *Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm.* Medical Imaging, IEEE Transactions on, 2001. **20**(1): p. 45-57.

104. Held, K., et al., *Markov random field segmentation of brain MR images.* Medical Imaging, IEEE Transactions on, 1997. **16**(6): p. 878-886.

105. Santago, P. and H.D. Gage, *Quantification of MR brain images by mixture density and partial volume modeling.* Medical Imaging, IEEE Transactions on, 1993. **12**(3): p. 566-574.

106. Tesař, L., et al., *Medical image analysis of 3D CT images based on extension of Haralick texture features.* Computerized Medical Imaging and Graphics, 2008. **32**(6): p. 513-520.

107. Van Leemput, K., et al., *Automated model-based bias field correction of MR images of the brain.* IEEE transactions on medical imaging, 1999. **18**(10): p. 885-896.

108. Kanungo, T., et al., *An efficient k-means clustering algorithm: Analysis and implementation.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. **24**(7): p. 881-892.

109. Nock, R. and F. Nielsen, *On weighting clustering.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006. **28**(8): p. 1223-1235.

110. Heimann, T. and H.-P. Meinzer, *Statistical shape models for 3D medical image segmentation: a review.* Medical image analysis, 2009. **13**(4): p. 543-563.

111. Rueckert, D., A.F. Frangi, and J.A. Schnabel, *Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration.* Medical Imaging, IEEE Transactions on, 2003. **22**(8): p. 1014-1025.

112. Shen, D. and C. Davatzikos, *An adaptive-focus deformable model using statistical and geometric information.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000. **22**(8): p. 906-913.

113. Tsai, A., et al., *A shape-based approach to the segmentation of medical imagery using level sets.* Medical Imaging, IEEE Transactions on, 2003. **22**(2): p. 137-154.

114. Leventon, M.E., W.E.L. Grimson, and O. Faugeras. *Statistical shape influence in geodesic active contours*. in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. 2000. IEEE.

115. Rousson, M. and N. Paragios, *Shape priors for level set representations*, in *Computer Vision—ECCV 2002*2002, Springer. p. 78-92.

116. 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.

117. Osher, S. and R. Fedkiw, *Level set methods and dynamic implicit surfaces*. Vol. 153. 2003: Springer Verlag.

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

119. Chan, T. and L. Vese, *An active contour model without edges.* Scale-Space Theories in Computer Vision, 1999. **1682**: p. 141-151.

120. Chan, T.F. and L.A. Vese. *A level set algorithm for minimizing the Mumford-Shah functional in image processing*. in *Proceedings of IEEE Workshop on Variational and Level Set Methods in Computer Vision*. 2001. IEEE.

121. Li, C.M., et al., *Minimization of region-scalable fitting energy for image segmentation.* Ieee Transactions on Image Processing, 2008. **17**(10): p. 1940-1949.

122. Malladi, R., J.A. Sethian, and B.C. Vemuri, *Shape Modeling with Front Propagation - a Level Set Approach.* IEEE transactions on pattern analysis and machine intelligence, 1995. **17**(2): p. 158-175.

123. Li, C., et al., *Distance regularized level set evolution and its application to image segmentation.* Image Processing, IEEE Transactions on, 2010. **19**(12): p. 3243-3254.

124. Rohlfing, T., et al., *Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains.* NeuroImage, 2004. **21**(4): p. 1428-1442.

125. Warfield, S.K., K.H. Zou, and W.M. Wells, *Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation.* IEEE transactions on medical imaging, 2004. **23**(7): p. 903-921.

126. Isgum, I., et al., *Multi-atlas-based segmentation with local decision fusion--application to cardiac and aortic segmentation in CT scans.* IEEE transactions on medical imaging, 2009. **28**(7): p. 1000-10.

127. Langerak, T.R., et al., *Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE).* Medical Imaging, IEEE Transactions on, 2010. **29**(12): p. 2000-2008.

128. Allen, W.M. and C. Cheacker, *Mastitis and somatic cell counts.* The Veterinary record, 2003. **152**(1): p. 28.

129. Kittler, J., et al., *On combining classifiers.* Ieee Transactions on Pattern Analysis and Machine Intelligence, 1998. **20**(3): p. 226-239.

130. Kearns, M.J. and L.G. Valiant, *Learning Boolean formulae or finite automata is as hard as factoring*1988: Harvard University, Center for Research in Computing Technology, Aiken Computation Laboratory.

131. Schapire, R.E., *The strength of weak learnability.* Machine learning, 1990. **5**(2): p. 197-227.

132. Freund, Y. and R. Schapire. *A desicion-theoretic generalization of on-line learning and an application to boosting*. in *Computational learning theory*. 1995. Springer.

133. Sabuncu, M.R., et al., *A Generative Model for Image Segmentation Based on Label Fusion.* IEEE transactions on medical imaging, 2010. **29**(10): p. 1714-1729.

134. Wang, H., et al., *Multi-atlas segmentation with joint label fusion.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013. **35**(3): p. 611-623.

135. Commowick, O. and S.K. Warfield, *Incorporating priors on expert performance parameters for segmentation validation and label fusion: a maximum a posteriori STAPLE.* Medical image

computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, 2010. **13**(Pt 3): p. 25-32.

136. Commowick, O., A. Akhondi-Asl, and S.K. Warfield, *Estimating A Reference Standard Segmentation With Spatially Varying Performance Parameters: Local MAP STAPLE.* IEEE transactions on medical imaging, 2012. **31**(8): p. 1593-1606.
137. Landman, B.A., et al., *Robust Statistical Fusion of Image Labels.* IEEE transactions on medical imaging, 2012. **31**(2): p. 512-522.
138. Asman, A.J. and B.A. Landman, *Robust Statistical Label Fusion Through Consensus Level, Labeler Accuracy, and Truth Estimation (COLLATE).* IEEE transactions on medical imaging, 2011. **30**(10): p. 1779-1794.
139. Asman, A.J. and B.A. Landman, *Formulating spatially varying performance in the statistical fusion framework.* IEEE transactions on medical imaging, 2012. **31**(6): p. 1326-36.
140. Asman, A.J. and B.A. Landman, *Non-local statistical label fusion for multi-atlas segmentation.* Medical image analysis, 2013. **17**(2): p. 194-208.
141. Cardoso, M.J., et al., *STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation.* Medical image analysis, 2013. **17**(6): p. 671-684.
142. Cardoso, M.J., et al., *LoAd: a locally adaptive cortical segmentation algorithm.* NeuroImage, 2011. **56**(3): p. 1386-97.
143. Song, Z., et al., *Integrated graph cuts for brain MRI segmentation*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*2006, Springer. p. 831-838.
144. Hu, S., E.A. Hoffman, and J.M. Reinhardt, *Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images.* IEEE transactions on medical imaging, 2001. **20**(6): p. 490-8.
145. Chen, C.W., J. Luo, and K.J. Parker, *Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications.* IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 1998. **7**(12): p. 1673-83.

146. Dice, L.R., *Measures of the amount of ecologic association between species.* Ecology, 1945. **26**(3): p. 297-302.

147. Poulose, B.K., et al., *Epidemiology and cost of ventral hernia repair: making the case for hernia research.* Hernia : the journal of hernias and abdominal wall surgery, 2012. **16**(2): p. 179-83.

148. Forbes, S.S., et al., *Meta-analysis of randomized controlled trials comparing open and laparoscopic ventral and incisional hernia repair with mesh.* British Journal of Surgery, 2009. **96**(8): p. 851-858.

149. Millikan, K.W., *Incisional hernia repair.* Surgical Clinics of North America, 2003. **83**(5): p. 1223-1234.

150. Breuing, K., et al., *Incisional ventral hernias: review of the literature and recommendations regarding the grading and technique of repair.* Surgery, 2010. **148**(3): p. 544-558.

151. Kanters, A.E., et al., *Modified hernia grading scale to stratify surgical site occurrence after open ventral hernia repairs.* Journal of the American College of Surgeons, 2012. **215**(6): p. 787-793.

152. Tanaka, E.Y., et al., *A computerized tomography scan method for calculating the hernia sac and abdominal cavity volume in complex large incisional hernia with loss of domain.* Hernia : the journal of hernias and abdominal wall surgery, 2010. **14**(1): p. 63-69.

153. McAuliffe, M.J., et al. *Medical image processing, analysis and visualization in clinical research*. in *Proceedings of the 14th IEEE Symposium on Computer-Based Medical Systems*. 2001. IEEE.

154. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**(1): p. 159-74.

155. Wahba, G. and Society for Industrial and Applied Mathematics., *Spline models for observational data*, in *CBMS-NSF Regional Conference series in applied mathematics 59*1990, Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104): Philadelphia, Pa.

156. Hosmer, D.W., S. Lemeshow, and R.X. Sturdivant, *Introduction to the logistic regression model*2000: Wiley Online Library.

157. Zou, H. and T. Hastie, *Regularization and variable selection via the elastic net.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005. **67**(2): p. 301-320.

158. Cristianini, N. and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*2000: Cambridge university press.

159. Ousley, J.B., Rebeccah B.; Holzman, Michael D.; Ehrenfeld, Jesse M.; Sharp, Kenneth W.; Nealon, William H.; Poulose, Benjamin K., *History of MRSA Infection Considerably Increases Risk of Surgical Site Infection in Ventral Hernia Repair*. Journal of the American College of Surgeons, 2014. **219**: p. S99.

160. Slater, N.J., et al., *Criteria for definition of a complex abdominal wall hernia.* Hernia : the journal of hernias and abdominal wall surgery, 2013: p. 1-11.

161. Pans, A., et al., *Long-term results of polyglactin mesh for the prevention of incisional hernias in obese patients.* World journal of surgery, 1998. **22**(5): p. 479-483.

162. Trimbos, J.B., et al., *A randomized clinical trial comparing two methods of fascia closure following midline laparotomy.* Archives of surgery, 1992. **127**(10): p. 1232-4.

163. Aguilar, B., et al., *Conservative Management of Mesh-Site Infection in Hernia Repair*. Journal of Laparoendoscopic & Advanced Surgical Techniques, 2010. **20**(3): p. 249-252.

164. Lo, C.-Y., et al. *DTI-based virtual reality system for neurosurgery*. in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. 2007. IEEE.

165. Gallo, L., *A glove-based interface for 3D medical image visualization*, in *Intelligent interactive multimedia systems and services*2010, Springer. p. 221-230.

166. Zhang, S., et al. *An immersive virtual environment for DT-MRI volume visualization applications: a case study*. in *Visualization, 2001. VIS'01. Proceedings*. 2001. IEEE.

167. Gallo, L., et al. *Toward a natural interface to virtual medical imaging environments*. in *Proceedings of the working conference on Advanced visual interfaces*. 2008. ACM.

168. Indhumathi, C., W. Chen, and Y. Cai, *Multi-modal VR for medical simulation.* The International Journal of Virtual Reality, 2009. **8**(1): p. 1-7.

169. O'Connor, J.P.B., et al., *Quantitative imaging biomarkers in the clinical development of targeted therapeutics: current and future perspectives.* The lancet oncology, 2008. **9**(8): p. 766-776.

170. Feng, D., L. Wee Kheng, and S. Venkatesh. *Removal of abdominal wall for 3D visualization and segmentation of organs in CT volume*. in *Proceedings of the 16th IEEE International Conference on Image Processing*. 2009.

171. Kansal, A.R., et al., *Cellular automaton of idealized brain tumor growth dynamics.* Bio Systems, 2000. **55**(1-3): p. 119-27.

172. Chellappa, R. and S. Chatterjee, *Classification of textures using Gaussian Markov random fields.* Acoustics, Speech and Signal Processing, IEEE Transactions on, 1985. **33**(4): p. 959-963.

173. Chen, C.-C., J.S. DaPonte, and M.D. Fox, *Fractal feature analysis and classification in medical imaging.* Medical Imaging, IEEE Transactions on, 1989. **8**(2): p. 133-142.

174. Paragios, N. and R. Deriche. *Geodesic active regions for supervised texture segmentation*. in *Proceedings of the 7th IEEE International Conference on Computer Vision*. 1999. IEEE.

175. Li, C., et al. *Fast distance preserving level set evolution for medical image segmentation*. in *Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision*. 2006. IEEE.

176. Qian, Z., D.N. Metaxas, and L. Axel. *Extraction and tracking of MRI tagging sheets using a 3D Gabor filter bank.* in *Proceedings of the IEEE 28th Annual International Conference on Engineering in Medicine and Biology Society*. 2006. IEEE.

177. Zhang, J., T. Tan, and L. Ma. *Invariant texture segmentation via circular Gabor filters*. in *Proceedings of the 16th International Conference on Pattern Recognition*. 2002. IEEE.

178. Smith, S.M. and J.M. Brady, *SUSAN—A new approach to low level image processing.* International Journal of Computer Vision, 1997. **23**(1): p. 45-78.

179. Jenkinson, M. and S. Smith, *A global optimisation method for robust affine registration of brain images.* Medical image analysis, 2001. **5**(2): p. 143-56.

180. Patenaude, B., et al., *A Bayesian model of shape and appearance for subcortical brain segmentation.* Neuroimage, 2011. **56**(3): p. 907-922.

181. Okada, T., et al., *Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors.* Medical image analysis, 2015. **26**(1): p. 1-18.

182. Pohl, K.M., et al., *Logarithm odds maps for shape representation*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*2006, Springer. p. 955-963.

183. Chan, T.F. and L. Vese, *Active contours without edges.* Image Processing, IEEE Transactions on, 2001. **10**(2): p. 266-277.

184. Iglesias, J.E. and M.R. Sabuncu, *Multi-Atlas Segmentation of Biomedical Images: A Survey.* arXiv preprint arXiv:1412.3421, 2014.

185. Xu, Z., et al. *Whole Abdominal Wall Segmentation using Augmented Active Shape Models (AASM) with Multi-Atlas Label Fusion and Level Set*. in *SPIE Medical Imaging*. 2016. San Diego, CA.

186. Cootes, T., E. Baldock, and J. Graham, *An introduction to active shape models.* Image processing and analysis, 2000: p. 223-248.

187. Criminisi, A., et al., *Regression forests for efficient anatomy detection and localization in computed tomography scans.* Medical image analysis, 2013. **17**(8): p. 1293-1303.

188. Modat, M., et al., *Fast free-form deformation using graphics processing units.* Computer methods and programs in biomedicine, 2010. **98**(3): p. 278-284.

189. Wheeler-Kingshott, C., et al., *The current state-of-the-art of spinal cord imaging: applications.* Neuroimage, 2014. **84**: p. 1082-1093.

190. Lycklama, G., et al., *Spinal-cord MRI in multiple sclerosis.* The Lancet Neurology, 2003. **2**(9): p. 555-562.

191. Smith, S.M., et al., *Advances in functional and structural MR image analysis and implementation as FSL.* Neuroimage, 2004. **23**: p. S208-S219.

192. Asman, A.J., et al., *Groupwise multi-atlas segmentation of the spinal cord's internal structure.* Medical image analysis, 2014. **18**(3): p. 460-471.

193. Shen, D., E.H. Herskovits, and C. Davatzikos, *An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures.* Medical Imaging, IEEE Transactions on, 2001. **20**(4): p. 257-270.

194. Li, B., et al. *Regression forest region recognition enhances multi-atlas spleen labeling*. in *MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA)*. 2013. Nagoya, Japan.

195. Asman, A.J. and B.A. Landman, *Non-local STAPLE: An intensity-driven multi-atlas rater model*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*2012, Springer. p. 426-434.

196. Okada, T., et al., *Abdominal multi-organ ct segmentation using organ correlation graph and prediction-based shape and location priors*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*2013, Springer. p. 275-282.

197. Avants, B.B., N. Tustison, and G. Song, *Advanced normalization tools (ANTS).* Insight J, 2009.

198. Heimann, T., et al., *Comparison and evaluation of methods for liver segmentation from CT datasets.* Medical Imaging, IEEE Transactions on, 2009. **28**(8): p. 1251-1265.

199. Gass, T., G. Szekely, and O. Goksel, *Multi-atlas segmentation and landmark localization in images with large field of view*, in *Medical Computer Vision: Algorithms for Big Data*2014, Springer. p. 171-180.

200. Kéchichian, R., et al., *Shortest-path constraints for 3d multiobject semiautomatic segmentation via clustering and graph cut.* Image Processing, IEEE Transactions on, 2013. **22**(11): p. 4224-4236.

201. Sullivan, D.C., *Imaging as a quantitative science*. Radiology, 2008. **248**(2): p. 328-32.

202. Parkinson, H., et al., *ArrayExpress update--an archive of microarray and high-throughput sequencing-based functional genomics experiments.* Nucleic Acids Res, 2011. **39**(Database issue): p. D1002-4.

203. Barrett, T., et al., *NCBI GEO: archive for high-throughput functional genomic data.* Nucleic Acids Res, 2009. **37**(Database issue): p. D885-90.

204. Bourgon, R., et al., *High-throughput detection of clinically relevant mutations in archived tumor samples by multiplexed PCR and next-generation sequencing.* Clin Cancer Res, 2014. **20**(8): p. 2080-91.

205. Roychowdhury, S., et al., *Personalized oncology through integrative high-throughput sequencing: a pilot study.* Sci Transl Med, 2011. **3**(111): p. 111ra121.

206. Wolf, A.D. and J.E. Lavine, *Hepatomegaly in neonates and children.* Pediatr Rev, 2000. **21**(9): p. 303-10.

207. O'Reilly, R.A., *Splenomegaly in 2,505 patients at a large university medical center from 1913 to 1995. 1963 to 1995: 449 patients.* West J Med, 1998. **169**(2): p. 88-97.

208. Felker, G.M., et al., *Underlying causes and long-term survival in patients with initially unexplained cardiomyopathy.* N Engl J Med, 2000. **342**(15): p. 1077-84.

209. Eisenhauer, E.A., et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).* Eur J Cancer, 2009. **45**(2): p. 228-47.

210. Dittrich, M., et al., *Sonographic biometry of liver and spleen size in childhood.* Pediatr Radiol, 1983. **13**(4): p. 206-11.

211. Bezerra, A.S., et al., *Determination of splenomegaly by CT: is there a place for a single measurement?* American Journal of Roentgenology, 2005. **184**(5): p. 1510-1513.

212. Overman, M.J., et al., *Oxaliplatin-mediated increase in spleen size as a biomarker for the development of hepatic sinusoidal injury.* J Clin Oncol, 2010. **28**(15): p. 2549-55.

213. Tefferi, A., et al., *Revised response criteria for myelofibrosis: International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consensus report.* Blood, 2013. **122**(8): p. 1395-8.

214. Rorden, C. and M. Brett, *Stereotaxic display of brain lesions.* Behavioural neurology, 2000. **12**(4): p. 191-200.

215. Harrison, C., et al., *JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis.* N Engl J Med, 2012. **366**(9): p. 787-98.

216. West, J., et al., *Comparison and evaluation of retrospective intermodality brain image registration techniques.* Journal of computer assisted tomography, 1997. **21**(4): p. 554-568.

217. Fitzpatrick, J.M. and J.B. West, *The distribution of target registration error in rigid-body point-based registration.* Medical Imaging, IEEE Transactions on, 2001. **20**(9): p. 917-927.

218. Murphy, K., et al., *Evaluation of methods for pulmonary image registration: The EMPIRE10 study*. Grand Challenges in Medical Image Analysis, 2010. **2010**.

219. Jenkinson, M., et al., *Fsl.* Neuroimage, 2012. **62**(2): p. 782-790.

220. Heinrich, M.P., et al., *MRF-based deformable registration and ventilation estimation of lung CT.* Medical Imaging, IEEE Transactions on, 2013. **32**(7): p. 1239-1248.

221. Heinrich, M.P., et al., *Towards realtime multimodal fusion for image-guided interventions using self-similarities*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*2013, Springer. p. 187-194.

222. Menke, J. and T.R. Martinez. *Using permutations instead of student's t distribution for p-values in paired-difference algorithm comparisons*. in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*. 2004. IEEE.

223. Bechhofer, R.E., *A single-sample multiple decision procedure for ranking means of normal populations with known variances.* The Annals of Mathematical Statistics, 1954: p. 16-39.

224. Vercauteren, T., et al., *Diffeomorphic demons: Efficient non-parametric image registration.* Neuroimage, 2009. **45**(1): p. S61-S72.

225. Urschler, M., et al., *Robust optical flow based deformable registration of thoracic CT images.* Medical Image Analysis for the Clinic: A Grand Challenge, 2010: p. 195-204.

226. Okada, T., et al., *Abdominal multi-organ segmentation of CT images based on hierarchical spatial modeling of organ interrelations*, in *Abdominal Imaging. Computational and Clinical Applications*2012, Springer. p. 173-180.

227. Zhou, X., et al., *Constructing a probabilistic model for automated liver region segmentation using non-contrast X-ray torso CT images*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*2006, Springer. p. 856-863.

228. del Toro, O.A.J. and H. Müller, *Multi-structure atlas-based segmentation using anatomical regions of interest*, in *Medical Computer Vision. Large Data in Medical Imaging*2014, Springer. p. 217-221.

229. Amir, A., et al., *AALIM: a cardiac clinical decision support system powered by advanced multi-modal analytics.* Studies in health technology and informatics, 2009. **160**(Pt 2): p. 846-850.

230. Asman, A.J., et al., *Multi-atlas learner fusion: An efficient segmentation approach for large-scale data.* Medical image analysis, 2015. **26**(1): p. 82-91.

231. Liu, J., et al. *Automated segmentation of the thyroid gland on CT using multi-atlas label fusion and random forest*. in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. 2015. IEEE.

232. Roth, H.R., et al., *Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications*, in *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*2015, Springer. p. 3-12.

233. Pratx, G. and L. Xing, *GPU computing in medical physics: A review.* Medical physics, 2011. **38**(5): p. 2685-2697.

234. Franz, A., et al. *Annotation-free probabilistic atlas learning for robust anatomy detection in CT images*. in *SPIE Medical Imaging*. 2015. International Society for Optics and Photonics.

235. Klinder, T., et al., *Automated model-based rib cage segmentation and labeling in CT images*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*2007, Springer. p. 195-202.

236. Wang, H., et al., *A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation.* Neuroimage, 2011. **55**(3): p. 968-985.

237. Wang, H., Y. Cao, and T.F. Syed-Mahmood. *Finding a Path for Segmentation Through Sequential Learning*. in *Information Processing in Medical Imaging*. 2015. Springer.

238. Roth, H.R., et al. *Deep convolutional networks for pancreas segmentation in CT imaging*. in *SPIE Medical Imaging*. 2015. International Society for Optics and Photonics.

239. Wang, C. and O. Smedby. *Automatic multi–organ segmentation using fast model based level set method and hierarchical shape priors*. in *Proceedings of the VISCERAL Challenge at ISBI, CEUR Workshop Proceedings*. 2014.