

12-2011

## Evaluation of deformable image registration for improved 4D CT-derived ventilation for image guided radiotherapy

Richard Castillo

Follow this and additional works at: [https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations](https://digitalcommons.library.tmc.edu/utgsbs_dissertations)



Part of the [Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons](#), [Medical Sciences Commons](#), and the [Other Physics Commons](#)

---

### Recommended Citation

Castillo, Richard, "Evaluation of deformable image registration for improved 4D CT-derived ventilation for image guided radiotherapy" (2011). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 188.  
[https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations/188](https://digitalcommons.library.tmc.edu/utgsbs_dissertations/188)

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact [digitalcommons@library.tmc.edu](mailto:digitalcommons@library.tmc.edu).

## Chapter 1

### Background and Introduction

#### 1.1 Motivation

Nearly all patients receiving thoracic radiotherapy develop some degree of lung injury, causing a loss of pulmonary function and death in some patients. Consequently, pulmonary injury is the dose limiting toxicity in thoracic radiotherapy for lung cancer, making local control (based on bronchoscopic biopsy) achievable in less than 20% of patients [1]. The standard radiation dose is inadequate in these patients and is one reason for their poor survival rates.

Image guided radiotherapy (IGRT) seeks to apply imaging modalities for improved tumor targeting and normal tissue avoidance in order to improve tumor control and reduce complications. In 1990, Abratt et al. showed that delivery of radiotherapy through hypo-perfused pulmonary regions for lung cancer treatment results in less pulmonary injury [1]. In that prospective study of lung cancer patients, planar perfusion imaging was utilized to estimate the fraction of perfusion in the upper, lower, and middle lung zones. The group of patients whose irradiated pulmonary region contained in excess of 35% of the total perfusion was significantly ( $p = 0.024$ ) more likely to suffer a loss of diffusion capacity and worsening of their dyspnea score. The study concluded that less lung damage can be expected if there is decreased perfusion in the pulmonary regions irradiated. This prospective study suggested a role for pulmonary function imaging to minimize the functional lung irradiated and reduce treatment-related pulmonary complications. Treatment planning studies have demonstrated that concept to reduce irradiation of functional lung regions using either 3D conformal [2, 3] or intensity modulated radiotherapy [2-5]. These findings suggest that the use of physiological images in radiotherapy treatment planning to quantify regions of functional avoidance may prove an effective IGRT strategy for minimizing the injury and/or functional loss following radiotherapy. However, pulmonary function imaging requires equipment not available in radiation oncology clinics.

#### 1.2 Ventilation Imaging

Presently, pulmonary ventilation is most often evaluated from planar projection images or single photon emission computed tomography (SPECT) imaging acquired either during or following administration of a radioactive tracer. Multiple tracers have been utilized for SPECT ventilation imaging, including radioactive gases such as  $^{81m}\text{Kr}$  and  $^{133}\text{Xe}$ , as well as particulate aerosols such as  $^{99m}\text{Tc}$  diethylenetriamine-pentaacetic acid (DTPA) and  $^{99m}\text{Tc}$  Technegas.  $^{99m}\text{Tc}$  DTPA, which is comparably inexpensive to produce and has obtained regulatory approval in the United States, is the most commonly implemented radiopharmaceutical for clinical ventilation imaging [2]. The

resulting images are constructed from the distribution of radioactivity that results from normal tidal breathing following administration. However, several studies have shown that the distribution of  $^{99m}\text{Tc}$  DTPA radioactivity on planar images differs from the corresponding distribution of ventilation depicted with  $^{81m}\text{Kr}$  [3-5]. Furthermore, central airway deposition of the radioactive aerosol is a common source of degradation in the resulting ventilation image, which may result in regional loss of the underlying pulmonary function information.

Xenon-enhanced computed tomography (CT), another method for ventilation imaging, is based on the enhancement of CT Hounsfield Units (HUs) by inhaled xenon gas. The degree of enhancement is linearly proportional to the regional concentration of inhaled xenon gas [6-8]. Quantitative ventilation images are constructed by fitting exponential models to either the wash-in or wash-out density time curves in designated regions of interest (ROIs) [9, 10]. Specific ventilation can be calculated for each 3D ROI from the observed regional time constant for the exponential model. Limitations to this approach include the relatively high cost and limited availability of the xenon gas. Additionally, the anesthetic effect of xenon at high concentrations precludes its routine implementation as a suitable contrast agent for ventilation imaging. Finally, the dynamic imaging approach requires acquisition of high temporal resolution time sequence images at a fixed axial coverage, which is limited to the range from about 2.5 to 12.5 cm per xenon-CT study [7].

Techniques in magnetic resonance (MR) imaging have also received attention recently for application in ventilation imaging. Edelman et al. first proposed the use of oxygen as a contrast agent for generating ventilation images from MR in 1996 [11]. With oxygen-enhanced MR, volumetric images obtained while breathing room air are subtracted from those obtained while breathing 100% oxygen to generate an image of the distribution of molecular oxygen within the lung parenchyma [12]. These images are intended to represent functional information, though the contribution of pulmonary ventilation to the resulting distribution is uncertain. Other MR-based approaches to ventilation imaging include the use of hyperpolarized gases such as  $^{129}\text{Xe}$  or  $^3\text{He}$  as paramagnetic contrast agents. Routine application of MR-based techniques is subject to numerous drawbacks, including the requirement for tracer gases and specialized equipment. Furthermore, the degree of hyperpolarization is time dependent and the derived images do not yield quantitative physiological values.

### **1.3 CT Ventilation**

In 2000, Simon proposed a methodology for extracting functional information from pairs of breath-hold CT images, based on the average CT values in manually registered sub-volumes [13].

Assuming that the density of any pulmonary sub-volume can be expressed as a linear combination of water- and air-like material, it can be shown that the fractional volume change can be expressed:

$$\frac{\Delta V}{V_{exhale}} = 1000 \frac{\bar{H}_{inhale}^{VOI} - HU_{exhale}}{HU_{exhale} (1000 + \bar{H}_{inhale}^{VOI})} \quad (1)$$

where  $\bar{H}_{inhale}^{VOI}$  is the average of all  $HU$  corresponding to the set of inhalation-state voxels that correspond physically with the underlying anatomy contained within the exhalation-state voxel with CT number  $HU_{exhale}$ . The quantity  $\Delta V = V_{inhale} - V_{exhale}$  is the regional volume change due to inspiration, and the ratio  $\Delta V/V_{exhale}$  is by definition the physiologic specific ventilation. In 2005, Guerrero et al. implemented this framework, utilizing a volumetric deformable image registration (DIR) calculation in order to determine an approximation to equation (1) on a voxel-wise basis for 22 exhale/inhale breath-hold CT image pairs [14]. In 2006, it was further demonstrated that consecutive application of the same approximation to equation(1) between the component phases of a 4D CT and a reference phase volume could yield temporally resolved dynamic ventilation images, in which the distribution of ventilation could be obtained at multiple respiratory phase increments [15]. One advantage of this CT-based functional imaging strategy is that, since the ventilation images can be derived from the treatment-planning 4D CT, no additional imaging sessions are required. Moreover, the ventilation images are inherently registered with a reference phase volume from the 4D CT, thus no additional multi-modality image alignment is necessary for incorporation of the functional information into radiotherapy treatment planning. Furthermore, the radioactive tracers and specialized equipment beyond what is currently implemented for standard thoracic radiotherapy treatment planning is not required. In 2007, Yaremko et al. demonstrated in a retrospective study that the ventilation images derived from treatment-planning 4D CT could be successfully incorporated into inverse planning for thoracic IMRT, resulting in improved functional dosimetry without compromising standard volumetric planning criteria [16]. Together, these data suggest that 4D CT-based ventilation imaging, which is more broadly available, can provide critical additional treatment planning information at the cost of additional computations only.

Ultimately, the benefits associated with the incorporation of the additional functional information into thoracic radiotherapy treatment planning will need to be tested in a prospective clinical trial to determine the impact on patient outcomes. However, validation of the 4D CT-



derived ventilation imaging technique with the current clinical standard SPECT ventilation imaging is necessary before clinical implementation can be realized. Currently, this validation step is lacking. Determining whether clinically-relevant functional information can be extracted from 4D CT images is a difficult task. In previous works mentioned above [14-16], preliminary evaluation has been based on the correlation of tidal volume calculations derived from the ventilation images with corresponding measurements obtained by manual segmentation of lung voxels from the CT image pairs. In both studies, linear correlation between measured and calculated tidal volumes was demonstrated, with corresponding  $R^2 = 0.97$  for comparison over 22 [14] and 3 [15] clinically acquired patient cases. However, regional assessment of the distribution of ventilation has not been performed.

DIR, the process of determining a spatial relationship between corresponding volume elements in volumetric images, is central to the ventilation calculation. Spatially accurate DIR is required to make the correspondence of HUs (equation(1)) between CT image pairs physically meaningful and allow the calculation of physiologically accurate ventilation images. Thus, spatial inaccuracies in the DIR transformation are expected to result in degradation of the resulting ventilation image, thereby contributing as a source of error to the functional information extracted from 4D CT. The incorporation of such errors into the ventilation images can lead to diminished correlation when comparison is made with standard ventilation imaging methods. Therefore, this study will first investigate the impact of DIR spatial accuracy on the ventilation image output. A methodology for evaluating the quality of DIR output for deformable registration of thoracic CT image pairs is required.

#### **1.4 DIR Evaluation**

DIR evaluation is currently an active area of our research, requiring careful consideration towards both the metric and the medium utilized to assess the characteristic qualities of a given image registration algorithm.

Synthetic images represent a commonly employed tool for assessing DIR performance [17-22]. Generally, a single clinically acquired volumetric image is obtained and subsequently modified according to a predefined transformation in order to generate a second, synthetic image, whose correspondence with the base image is known precisely at each location. The complexity of the applied deformation can range from simple three-dimensional affine transformations [17], to more complex simulations based on biomechanical models [18]. Additionally, purely synthetic deforming objects have also been utilized to assess DIR. Computer generated objects are especially useful for evaluating the robustness of DIR formulations as intrinsic model

assumptions are made to be incrementally violated by the input image content. Computer simulations and synthetic image pairs offer the obvious advantage of having a known correspondence at each image location. Thus, objective evaluation of DIR can be performed over the entire image domain. In addition, model parameters typically allow control over image variables such as noise type, level and distribution, as well as simulated physiological variables such as tumor volume or the presence of bowel gas.

Mechanical phantoms offer an alternative to computer simulation, and provide authentic image acquisitions of physical objects in motion as reference for evaluation of DIR [20, 23-25]. Drive assemblies can be used to precisely control the displacement of the phantom during image acquisition, with programmable motion patterns typically allowing for variable sinusoidal trajectories of a rigid object along three orthogonal axes. For evaluation of thoracic DIR, more complex deforming phantoms have also been reported. For example, Kashani et al. [26] reported the use of a diagnostic thoracic phantom, modified to include an actuator-driven lucite diaphragm, which in turn applied variable compression to an iodine-infused foam lung insert to mimic the cyclical compression of the lung parenchyma during respiration. To facilitate DIR spatial accuracy evaluation, 48 plastic markers were inserted into the lung foam and subsequently delineated in two images of the phantom at differing compression states [23]. Mechanical phantoms are advantageous in the sense that, if acquired using consistent protocols, the phantom images are susceptible to the same image acquisition and reconstruction artifacts as are analogously acquired patient images. This realism is preferred over simulation of the same effects in synthetic images, though simulation provides for far more complexity in the geometric transformation. Similarly to synthetic image pairs, for rigid body phantoms the correspondence between images is dictated by the applied motion, and is thus easily inferred at each image location. For deforming phantoms, however, the correspondence between images is not straightforward, and is thus limited to a small set of implanted markers.

The fundamental difficulty associated with objective evaluation of DIR for use in the clinical setting is due to the lack of "ground truth," or a known solution against which the registration output can be objectively compared. Simulation and mechanical phantoms provide a means for assessing DIR that is based on the deliberate construction of a known transformation between images. This type of assessment can be useful for model verification during development phases, in which it is desirable to explore the impact of various isolated variables on the DIR output [27]. However, simulations and mechanical phantoms lack sufficient realism to provide credible indication of DIR performance in clinical application. The confounding factor is the combination of image acquisition and reconstruction artifacts encountered in practice, the

complex physiological motion of shifting and deforming anatomic structures, and physiologically-induced changes in image content, such as the effect of regional gas volume on the CT HU within the lung. In addition, simulations have the added drawback of introducing potential bias towards those algorithms that are based upon models or formulations similar to the artificial deforming mechanism, thus limiting their applicability when evaluating DIR for clinical use. Deformable phantoms represent an attempt to bridge the gap between clinically acquired image content and a surrogate for quantitative evaluation of DIR. However, currently it is necessary to manually identify implanted markers between corresponding image pairs for spatial accuracy assessment. Thus, it is not clear if there is any advantage to utilizing a deforming phantom, provided that reference landmark pairs could be analogously identified between clinically acquired images. Model verification and optimization notwithstanding, for an algorithm ultimately to be implemented in clinical application, thorough performance evaluation should be conducted utilizing image data that will be encountered in practice. The best standard, therefore, is one derived from actual patient images, for which ground truth is not known.

Many studies exist in the literature in which evaluation of an image registration algorithm is either performed out-right, or as a validation step for the image registration component of a larger application, using clinically acquired images [28-39]. However, the particular metric determined to assess DIR performance is still variable. Generally, a performance metric can be considered as any computable measure used for quantitatively evaluating the performance of an algorithm [40]. Three main classes of computable measures have typically been utilized for evaluation of deformable image registration; image similarity, principles of continuum mechanics, and spatial accuracy.

Image similarity-based metrics are the simplest to determine, and potentially can provide indication of regional gross mis-registration between images. These metrics are based on both visual and quantitative comparison of attributes pertaining to a designated base image volume, with the corresponding attributes determined according to the application of the calculated DIR transformation. For example, the correlation coefficient has commonly been employed to quantify the gray-value similarity between image pairs, either as a whole, or on a volumetric slice-by-slice basis, before and after deformable registration to demonstrate improvement of image similarity following DIR [37]. It has also been used as a measure of the relative performance of multiple algorithms [38]. Similarly, subtraction images, constructed by taking either the signed or absolute difference between images before and after DIR, have also been used for qualitative visual assessment of regional gross mis-registration, which is inferred from large intensity differences between registered volume elements [7]. Image similarity measures for

designated two- and three-dimensional regions of interest (ROI) are also commonly implemented for targeted assessment of DIR. Quantitative ROI-based metrics include the volume of overlap index between an initial ROI and its DIR-propagated component [39]. Similarly, qualitative measures include visual assessment of the concordance between a manually-drawn reference set of two-dimensional contours and the corresponding propagated set determined from DIR [28, 37, 39].

In a 2007 study, Zhong et al. reported on a novel finite-element based framework for automated assessment of deformable image registration, derived from considerations in continuum mechanics [41]. The concept of unbalanced energy is introduced as a metric for DIR performance that can be quantified at both anatomic and voxel levels. Visualization of the performance metric subsequently provides for rapid localization of erroneous regions of the calculated transformation. The authors propose that this methodology has application in both quantitative comparative evaluation of multiple algorithms, as well as routine quality assessment of individual DIR outputs. However, the methodology is relatively new and untested, and bears notable drawbacks. For example, the current formulation assumes material homogeneity, which is generally inaccurate at tissue interfaces, and raises concerns for the applicability of such a method for evaluation of thoracic DIR. Additionally, it is not clear that there is a consistent practical interpretation of the quantitative unbalanced energy for assessment of DIR performance characteristics. That is, it may be difficult to evaluate the usefulness of a given DIR output in terms of clinically relevant tolerances that are defined in metric quantities. Moreover, visualization of the quantitative unbalanced energy at best only provides for spatial correlation of relative error. Thus, the proposed framework does not provide for more detailed characterization of DIR output in terms of clinically relevant variables that could potentially affect DIR output such as tissue displacement magnitudes or motion characteristics.

Finally, the third class of computable measures for evaluation of deformable image registration is based on quantifying the spatial accuracy of calculated displacements of individual volume elements between image pairs. Measurements of position represent the natural units for quantifying mis-registration, since the DIR calculation is used to determine a spatial transformation between images. In addition, units of spatial accuracy are most relevant for determining clinical utility, as acceptable limits for uncertainties in most clinical applications are reported in terms of spatial tolerances. It is important to note that neither image similarity- nor continuum mechanics-based performance metrics provide explicit indication of global or regional spatial accuracy. Often, a direct relationship is erroneously assumed, as is the case when difference images and/or the quantitative gray-value similarity are utilized as metrics for relative

performance between algorithms. Recent publications for evaluation of deformable image registration reflect research interest of the medical imaging community at large towards evaluation methods grounded in assessment of spatial accuracy [29-34].

However, there is still a lack of a standard or uniform method for assessing DIR spatial accuracy performance. Without a reference transformation, spatial accuracy evaluation is limited to a set of voxel positions for which some estimate of the correspondence between images is available. Thus, the evaluation process becomes inherently statistical in nature, requiring large and spatially distributed samples of voxel correspondences. There is also the need to characterize the uncertainty in the reference data itself, in order to determine an unbiased estimate of global DIR performance with a narrow range of uncertainty. Most often the voxel correspondences are determined manually by an imaging expert in the relevant anatomical site, though an alternative approach based on hyperpolarized  $^3\text{He}$  magnetic resonance tagging was proposed in 2007 for evaluation of thoracic DIR [42]. The thoracic MR approach, however, is limited by a number of drawbacks including minimum spacing requirements for tagged grid elements due to diffusion of the  $^3\text{He}$  gas, as well as the necessity for fast MR imaging techniques for dynamic image acquisition. Most notably, the applicability of reference data composed explicitly of MR images is limited for evaluation of deformable image registration. For example, the performance characteristics of an algorithm as evaluated using an explicitly-MR database are not immediately relevant for decisions regarding clinical applicability for use with radiation therapy treatment planning 4DCT.

Thus, the manual identification of voxel correspondences, which is not strictly limited to a particular imaging modality or acquisition process, is a more applicable and relevant method. Nevertheless, there does not currently exist a common framework for quantitative evaluation of DIR spatial accuracy using manually identified voxel correspondences. As a result, investigators have utilized a variety of reference sample sizes, with equally varying spatial distributions for evaluation of DIR. This inconsistency in evaluation standards makes objective comparison of published DIR spatial accuracy measurements difficult and potentially misleading. To eliminate these uncertainties, and to facilitate more detailed performance assessment of our 4D CT-based ventilation imaging procedure, we propose to develop a framework for objective evaluation of DIR spatial accuracy for deformable registration of thoracic CT images. For our purposes, the objective requirements on such a framework are two-fold, providing for both detailed evaluation and overall characterization of DIR spatial accuracy performance, as well as quality assurance checks for spatial accuracy assessment in routine clinical application. Such innovations will provide mathematical and statistical basis to guide researchers in the design of future DIR studies.

These tools will allow us to focus on the main problem of this proposal, extraction of functional information from 4D CT images and their associated deformations with attention to the impact of spatial registration inaccuracies on the fidelity of the associated output. In this study, evaluation of the 4D CT-derived ventilation images will aim to assess both the correlation with global measurements of lung ventilation, as well as regional correlation of the distribution of ventilation with the clinical reference standard  $^{99m}\text{Tc}$  DTPA single photon emission computed tomography (SPECT) imaging. The aim of this work is to assess correlation with the clinical references for lung function imaging, and to assess the variability that arises due to spatial inaccuracies in the DIR transformation central to the ventilation calculation. A novel framework for both the detailed assessment of an algorithm's DIR spatial accuracy performance characteristics as well as quality assurance checks for spatial accuracy assessment in routine application is needed. Indeed, these evaluation strategies will ultimately comprise integral components of a CT-based ventilation imaging framework, which will be central to future prospective clinical trials for functional avoidance in thoracic radiotherapy. Additionally, a DIR evaluation infrastructure will find utility beyond the present context, as technologies for determining the DIR spatial transformation and novel applications for utilizing them are continually presented in the medical imaging literature. The translation and routine implementation of these technologies into clinical application will require an established methodology for validation and performance evaluation of DIR that is currently lacking. Within the present context, rigorous assessment of the impact of variable DIR output on the quality of 4D CT-derived ventilation images relative to a clinical ventilation imaging reference will provide critical information for establishing guidelines and minimum performance standards for the DIR component used in 4D CT ventilation image calculations.

## **1.5 Innovation**

There is increasing interest in utilizing physiological images in radiotherapy treatment planning for image guidance to avoid the irradiation of highly functional regions and minimize the functional consequence of pulmonary injury following radiotherapy [43-47]. Currently the volume of normal lung that will receive a specified radiation dose is utilized in radiotherapy treatment planning to estimate the risk of pulmonary injury [48, 49]. This method assumes homogenous pulmonary function throughout the entire lung [50]. However, patients may have significantly heterogeneous regional lung function resulting from underlying physiological differences, chronic pulmonary disease, or malignancy [51-53]. A more accurate method of risk assessment would include both the volume and the functional capacity of the portions of the lung

intended for irradiation [44, 47]. Existing ventilation imaging techniques require a tracer gas and specialized imaging equipment such as an MR scanner or SPECT camera, and their resulting images do not provide quantitative physiological parameters. Furthermore, such equipment is not available in radiation oncology clinics. Studies utilizing these techniques in patients who receive thoracic irradiation have been limited to the Netherlands Cancer Institute, Princess Margaret Hospital, and Duke University. Each of those institutions has evaluated regional lung function response to irradiation using SPECT perfusion. A broadly available pulmonary function imaging modality for image-guided radiotherapy is lacking.

## **1.6 Scientific Contribution**

A consistent framework for objective evaluation of deformable image registration spatial accuracy performance is required for clinical evaluation of CT-based lung function imaging. Presently, such a framework is lacking. We will present a framework for objective evaluation of thoracic deformable image registration spatial accuracy, based on the use of large samples of expert-determined landmark feature pairs between volumetric images as a reference for spatial accuracy measurements. The proposed methodology will facilitate rigorous evaluation of DIR spatial accuracy performance, and find utility beyond the present context, providing mathematical and statistical basis to guide researchers in the design of future DIR studies.

Validation of 4D CT-derived ventilation imaging with nuclear medicine SPECT ventilation should be performed prior to translation into clinical implementation and routine use. No preliminary validation studies have been performed comparing 4D CT-derived ventilation with the clinical standard SPECT ventilation. Neither has any formal comparative evaluation of CT-derived ventilation with clinically acquired nuclear medicine SPECT perfusion been performed. These formal comparative evaluation studies versus the clinical reference lung function imaging techniques will provide for future studies to investigate the potential impact of CT-derived lung function information on radiotherapy treatment planning and patient outcomes in the treatment of thoracic malignancies.

## **1.7 Hypothesis and Objectives**

The hypothesis for this research is:

*4D CT-derived ventilation imaging will provide an accurate assessment of pulmonary function for use in image guided thoracic radiation therapy.*

The specific aims for testing the hypothesis are as follows:

**Specific Aim 1:** Establish a methodology for quantitative assessment and comparative evaluation of deformable image registration spatial accuracy in the lungs

**Specific Aim 2:** Measure the correlation between 4D CT-derived ventilation and the clinical standard  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation (using SPECT-CT)

**Specific Aim 3:** Evaluate the correlation between the pulmonary perfusion and ventilation defects for both ventilation imaging methods described in Specific Aim 2

## **1.8 Dissertation Organization**

Chapters 2 through 4 are each self-contained studies, including a formal introduction, materials and methods, results, discussion, and conclusion. These studies describe the work required to address each of the specific aims defined to test the hypothesis of this work. Specifically, Chapter 2 addresses Specific Aim 1, describing a novel framework for quantitative assessment of deformable image registration spatial accuracy performance, with application to comparative evaluation studies, and routine clinical quality assurance practices. Specific Aim 2 is addressed in Chapter 3, in which the distribution of ventilation derived from 4D CT is compared with the clinical reference nuclear medicine SPECT ventilation for a population of lung and esophagus cancer patients treated in the Department of Radiation Oncology at The University of Texas M. D. Anderson Cancer Center. Chapter 4 addresses Specific Aim 3 by performing similar spatial correlation analysis between 4D CT ventilation and nuclear medicine SPECT pulmonary perfusion imaging in a patient population with malignant airway stenosis. Chapter 5 describes an extension to the ventilation formalism provided in Chapters 3 and 4, in which we describe a methodology for explicitly determining the distinct contributions of pulmonary ventilation and perfusion to observed CT image intensities. Chapter 6 provides a general summary of the research project, as well as an evaluation of the hypothesis with respect to each of the Specific Aims. Also included in Chapter 6 is a description of multiple ongoing and future research applications derived from the present work. Finally, Chapter 7 is the Appendix, in which explicit derivation of the CT lung function equations from basic principles is provided.



## Chapter 2

### A Framework for Evaluation of Deformable Image Registration Spatial Accuracy using Large Landmark Point Sets

#### Preface

The following material included in this Chapter was originally published in the scientific literature prior to publication of this dissertation document. The original research was conducted as part of this dissertation research, and so is included here as it appears in the published manuscript. The citation for all indented ("[...]") material included in this Chapter is given:

[Richard Castillo, Edward Castillo, Rudy Guerra, Valen Johnson, Travis McPhail, Amit K. Garg, and Thomas Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. \*Physics in Medicine and Biology\*, 54: 1849-1870, 2009.](#)

#### 2.1 Introduction

" [The translation of image processing research into clinical application is generally confounded by the lack of established methodology for validating new algorithms [54, 55]. For objective and clinically relevant evaluation, considerable attention must be paid regarding the selection of an appropriate reference standard upon which to base algorithm performance and determine clinical utility. For validation of deformable image registration (DIR), a number of reference standards have been utilized, including synthetically deformed images [20, 22, 25], high-contrast phantoms [25], and expert-delineated control points [29, 30, 36, 56, 57]. While synthetic images and phantoms might provide useful qualitative evaluation of DIR performance characteristics, they lack sufficient realism to provide credible validation of registration spatial accuracy for use in the clinical setting [27]. The best standard, therefore, is one derived from actual patient image data, for which ground-truth is not known.

The relative abundance of high-contrast, anatomical landmarks such as vessel and bronchial bifurcations make thoracic 4D CT image data particularly well-suited for the manual tracking of prominent image features across multiple image volumes. Tracking such features offers a means for estimating the true transformation and provides measures for statistical analysis of DIR spatial accuracy. Recent published landmark-based validation studies of thoracic DIR reflect this notion [29-34, 57-61]. To date, however, there is not a common standard or framework for either generating or utilizing the reference samples used to characterize DIR performance. As a result, a large range of

reference sample sizes, with equally varying spatial distributions, have been used to validate novel DIR algorithms. This inconsistency in evaluation standards complicates the interpretation of individual validation studies, and makes objective comparison of reported DIR spatial accuracies difficult and potentially misleading.

For thorough and unbiased characterization of DIR spatial accuracy performance, it is necessary to ensure that the validation landmark sets adequately sample the volume of interest not only spatially, but also in terms of the clinically relevant variables that could potentially affect DIR output. Such factors include physiological motion characteristics such as displacement magnitude and hysteresis, image quality, and intensity characteristics such as local contrast and change in intensity between images. With this in mind, it is important to distinguish between quantitative assessment for characterization or acceptance testing, as opposed to quality assurance (QA) purposes. In the former, the goal is to construct a complete description of the DIR performance characteristics, and in the case of acceptance testing, utilizing as much of the available information as necessary to provide an informed assessment regarding the routine clinical feasibility and potential shortcomings of a given DIR algorithm. Thus, landmark samples should be selected of sufficient size to facilitate statistical analysis of DIR spatial accuracy performance. Though the selection of the necessarily large validation landmark sets is not feasible for routine QA purposes, it is also presumably not necessary, provided a thorough evaluation of the algorithm was performed prior to routine clinical implementation. For QA, the goal is rather to ensure for any given case that the DIR spatial accuracy meets accepted standards within the context of the specific clinical application. The insight acquired during the characterization process is therefore crucial, and directly applicable to the development of specific QA testing procedures, that, based on only a limited amount of information for any given case, will nonetheless ultimately be used to judge the quality of the output in order to prevent potentially harmful errors from reaching the patient.

The goal of this Aim is to demonstrate a proposed consistent and self-contained framework for objective performance evaluation of thoracic deformable image registration. This framework is based on the use of large samples of expert-determined landmark feature pairs between volumetric images as a reference for spatial accuracy measurements, for purposes of:

- optimization and characterization of DIR output during algorithm development,
- comparative evaluation of multiple DIR algorithms,

- formal acceptance testing of individual algorithms for specific clinical application, and
- quality assurance of DIR output in the routine clinical setting.

For landmark selection, we employ a novel MATLAB-based (Mathworks, Sunnyvale, CA) software interface, developed to streamline the manual selection process and manage the corresponding samples of validation point sets. Using the interface, large samples (>1100) of corresponding pulmonary landmark features were manually generated from treatment planning 4D CT data to facilitate statistical evaluation of DIR spatial accuracy. In order to demonstrate the practical utility of the landmark sets for validation and comparative evaluation, we compare the spatial accuracy performance of two DIR algorithms, a gradient-based optical flow algorithm and a landmark interpolation algorithm based on moving least squares, for registration of thoracic CT image pairs. Furthermore, we investigate the correlation of standard image intensity-based measures for assessing DIR performance with the spatial accuracy derived from the validation landmark sets. Finally, we utilize the statistical properties of the DIR output over the validation point sets to demonstrate the effect of landmark sample size on the uncertainty associated with calculated values for mean registration error.

The organization of the remainder of this chapter is as follows. Section 2.2 describes the process of generating the large samples of manually registered feature points for objective evaluation of DIR spatial accuracy. It is broken into three sections. Section 2.2.1 briefly describes the five clinically acquired patient images utilized throughout this study. Section 2.2.2 describes the experimental methods for landmark registration and section 2.2.3 describes the large landmark datasets generated from the five clinically acquired treatment planning 4D CT image volumes. Section 2.2.4 describes the statistical characterization of the landmark sets for uncertainties associated with observer variance, while section 2.2.5 addresses the issue of landmark localization uncertainty with regard to image resolution and voxel dimension. Section 2.3 demonstrates the practical utility and necessity of the large point sets both for characterization as well as comparative evaluation of DIR outputs. Sections 2.3.1 and 2.3.2 briefly describe the two DIR algorithms that are used in this study to generate example DIR datasets for the five patient cases. Section 2.3.3 focuses on the spatial accuracy characterization of both output sets derived from the validation landmarks. Additionally, in section 2.3.3 we investigate the correlation of standard image-intensity based measures for assessing DIR performance with the spatial accuracy measurements

derived from the validation landmark sets. Section 2.4 focuses on the statistical requirements on landmark sample size, with an example presented in the context of comparative evaluation of multiple DIR algorithms. Finally, Section 2.5 summarizes the framework and provides a general discussion regarding its use.

## **2.2 Reference Landmark Selection and Characterization**

While it has been demonstrated in the literature that manually registered landmark features can be a useful tool for obtaining spatial accuracy measurements for DIR, a consistent framework for evaluation is still lacking. This section describes a new methodology for generation of the manually registered point sets. For demonstration purposes, five clinically acquired patient data sets were initially obtained.

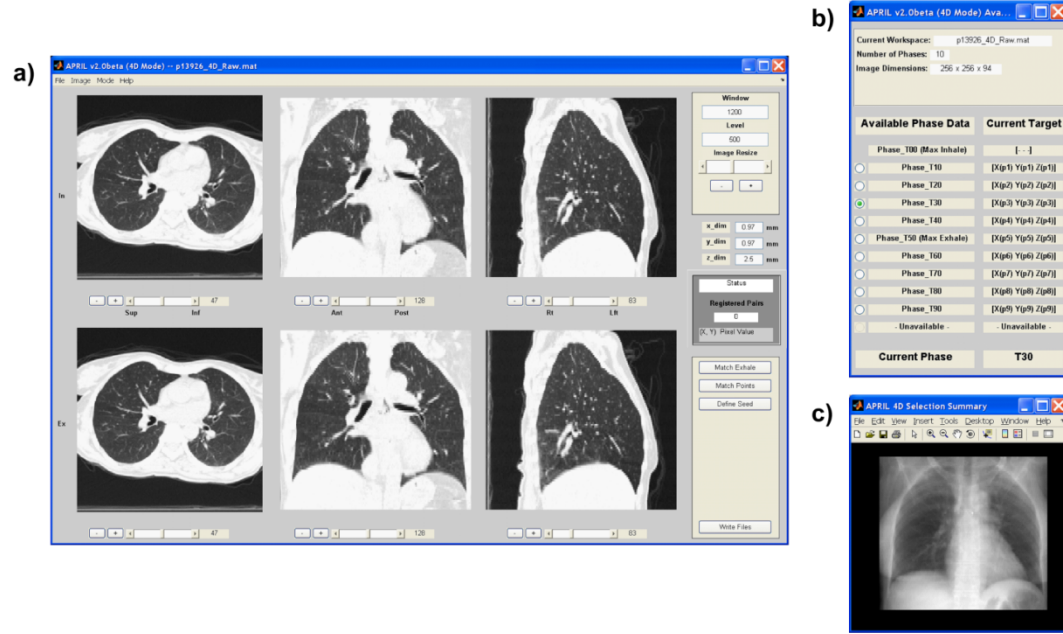
### *2.2.1 Thoracic CT Patient Images*

The treatment planning 4D CT images from five patients free of pulmonary disease who were treated for esophageal cancer were selected. Patient identifiers were removed in accordance with an institutional review board approved retrospective study protocol (RCR 03-0800). Each patient underwent treatment planning in which 4D CT images of the entire thorax and upper abdomen were acquired at 2.5 mm slice spacing with a General Electric Discovery ST PET/CT scanner (GE Medical Systems, Waukesha, WI). The extreme inhale and exhale phases of the 4D CT image sets were utilized in this study. Each image was cropped to include the entire rib cage and content sub-sampled to  $256 \times 256$  voxels. Final in-plane voxel dimensions ranged from  $(0.97 \times 0.97)$  to  $(1.16 \times 1.16)$  mm<sup>2</sup>. No sub-sampling was performed in the superior-inferior direction. For all cases, the final image slice thickness was 2.5 mm.

### *2.2.2 Reference Feature Selection*

A MATLAB-based software interface named APRIL (Assisted Point Registration of Internal Landmarks) was developed to facilitate manual selection of landmark feature pairs between multiple volumetric images. Up to 11 three-dimensional image matrices can be stored and loaded for display in a single APRIL workspace. This provides for storage of a typical 4D CT data set, comprised of 10 component 3D volumes, as well as for an additional supplementary volume, such as average or maximum intensity projection CT. In 4D mode, images are assumed to have equal volume dimensions.

Right-left (RL), anterior-posterior (AP), and superior-inferior (SI) voxel dimensions can be defined by the user to interactively adjust the aspect ratio of the image displays.



**Figure 2.1. APRIL registration interface.** **a)** Primary display window, showing designated source (T00) and target (T30) component volumes from a 4D CT set. The top panel of orthogonal displays is fixed to show the designated source volume, while the bottom panel shows the active target volume. Tools located directly beneath each image allow the orthogonal displays to be independently scrolled, panned, and/or magnified. **b)** Available phase data window, in which the user designates the active target volume from a list of the available images stored in the current workspace. When a target point is selected, its corresponding (RL, AP, SI) coordinates are displayed next to the respective target volume label. **c)** Selection summary window, displays an AP projection of the source image data. The display is updated as source points are selected, to show a projection of the current set of feature points for reference.

Figure 2.1 shows a screen capture of the main interface display. Two volumes are simultaneously displayed in transverse, coronal, and sagittal orientation in the primary display window (figure 2.1a). Separate window and level edit boxes allow the user to apply global window and level display settings to the set of image displays. The top panel is fixed to display the designated source volume, while the bottom panel will display whichever available phase is designated as the active target volume in the available phase data window (figure 2.1b). The active target volume is designated by selection of the

appropriate radio toggle, which causes the bottom image panels to be actively updated. Each individual image display can be scrolled, magnified, and/or panned independently of the other five displays. Current pan orientation and magnification setting are preserved during slice scrolling.

When the cursor is positioned over any of the images a pixel information tool on the right side of the main display window provides the corresponding pixel intensity and its current  $(x, y)$  position relative to the display orientation. The coordinate and intensity information are updated in real time as the cursor changes position. Alignment pushbuttons allow any arbitrary voxel to be visualized in each of the display orientations. When the alignment tool is enacted, a voxel is selected via mouse click, and its equivalent location highlighted in the remaining displays. The software also provides for cine loop playback of the available phase data in axial, coronal, or sagittal section. Any slice can be viewed through its cine sequence which, in turn, can be exported as an AVI file.

The manual registration process begins with the selection of a unique feature point within the designated source volume via mouse click on any of the orthogonal displays. Upon selection, the feature voxel is highlighted for reference in each of the source and target images. To assist the manual selection process for any single feature point, the software provides an optional feature localization tool based on normalized cross-correlation of a size-adjustable local voxel neighborhood [62, 63]. Given the feature voxel  $v$  located at position  $(S_x, S_y, S_z)$  in the source image, we create a  $(m \times m \times m)$  source neighborhood  $N_s$ , centered on  $v$ . The isotropic neighborhood dimension is given by  $m = 2\alpha + 1$ , where the parameter  $\alpha$  is chosen by the user from a list of available dimensions that range from 4-64 voxels. A user-defined intensity threshold is applied to  $N_s$  to generate a binary mask of the original source neighborhood, designated  $N_s^*$ . A local neighborhood of the same dimension is similarly defined centered on  $v$  in the target volume and the user threshold applied to generate  $N_t^*$ . The normalized cross-correlation coefficient  $\delta$  at position  $(x, y, z)$  in the target image is then given according to the following:

$$\delta(x, y, z) = \frac{\sum_{r=-\alpha}^{\alpha} \sum_{s=-\alpha}^{\alpha} \sum_{t=-\alpha}^{\alpha} [N_s^*(S_x + r, S_y + s, S_z + t) - \bar{N}_s^*] [N_T^*(x + r, y + s, z + t) - \bar{N}_T^*]}{\left\{ \sum_{r=-\alpha}^{\alpha} \sum_{s=-\alpha}^{\alpha} \sum_{t=-\alpha}^{\alpha} [N_s^*(S_x + r, S_y + s, S_z + t) - \bar{N}_s^*]^2 \sum_{r=-\alpha}^{\alpha} \sum_{s=-\alpha}^{\alpha} \sum_{t=-\alpha}^{\alpha} [N_T^*(x + r, y + s, z + t) - \bar{N}_T^*]^2 \right\}^{0.5}}, \quad (2)$$

where  $\bar{N}_s^*$  and  $\bar{N}_T^*$  are the average intensity values within the respective binary source and target neighborhood masks. To facilitate rapid feature localization,  $\delta$  is only calculated over the region shared by  $N_s$  and the target volume. By using only the binary source and target neighborhoods, cross-correlation is performed only over the local structural content, where the level of included structural detail is controlled via the user-defined intensity threshold. The target voxel representing the maximum of the 3D correlation function is highlighted in the target displays and represents an estimate of the feature correspondence. In practice, multiple correlations varying both neighborhood dimensions and/or intensity threshold may be performed. However, the user ultimately must manually designate the feature correspondence via mouse click on the target image. Following confirmation of the target selection, the process is repeated until the desired sample size and uniformity of distribution have been achieved.

In addition to the cross-correlation assistance tool, which is designed to assist the independent search of any single image feature, a second optional computer assistance tool is provided to facilitate rapid registration of larger quantities of features in a single, accelerated workflow. In this case, computer assistance is provided for a given source feature position by moving least squares interpolation, using the set of previously registered feature pairs to interpolate an estimate of the current displacement. The contribution of previously registered landmark pairs to the interpolated displacement is weighted inversely according to magnitude distance from the current three-dimensional source position. As more coordinate pairs are registered, more information becomes available for input into the interpolation assistance scheme, ideally further guiding the estimated displacement of each successive source feature to be registered.

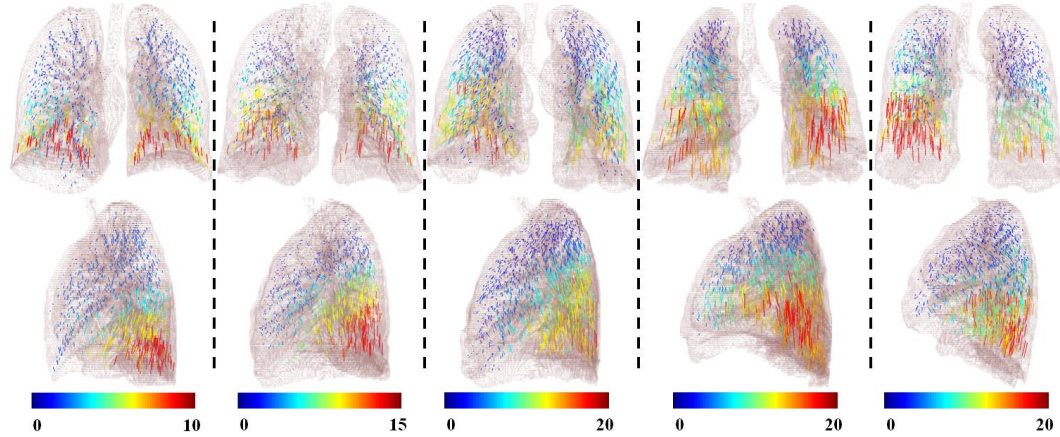
Following manual registration, summary text files are exported to streamline analysis procedures. The summary includes Cartesian and spherical coordinate lists of the source and target feature locations, corresponding voxel intensities, and displacement magnitudes (separately, in units of millimeters and voxels).

### 2.2.3 *Reference Landmark Datasets*

Pulmonary landmark feature pairs, typically vessel bifurcations, were manually delineated on the five test image pairs by an expert in thoracic imaging. Source feature points were selected systematically, beginning at the apex of the lung, with an initial goal of >10 feature points for each lung per axial image slice. This approach ensured the collection of >1100 validation point pairs for each case. Following feature selection for a given case, all landmark pairs were visually reviewed by the primary reader a second time and the location adjusted on the exhale image as necessary. The verification step was required before the initial registration process performed by the primary reader was considered complete.

The number of registered feature pairs per case ranged from 1166 to 1561. A total of 6762 landmarks were manually registered over the set of five image pairs. On average, approximately 12 hours, distributed over multiple sessions, were required to register a single case. Characteristics of the landmark pairs are summarized in Table 2.1. Average displacement and (standard deviation) of registered features per case ranged from 4.01 (2.91) – 9.42 (4.81) mm, while maximum landmark displacements ranged from 12.65 – 24.78 mm. Average magnitude displacements in component right-left (RL), anterior-posterior (AP), and superior-inferior (SI) directions ranged from 0.58 (0.62) – 1.17 (1.05) mm, 0.67 (0.79) – 1.74 (1.67) mm, and 3.68 (3.03) – 8.98 (5.04) mm, respectively. Figure 2.2 shows vector plots of the landmark displacement fields for the five cases in anterior (top row) and lateral (bottom row) projection. The sampled feature points are sufficiently distributed to illustrate substantially heterogeneous spatial distributions of tissue motion within each of the lung volumes.





**Figure 2.2. Reference landmark sets.** Manually determined displacement vectors are shown in anterior (top row) and lateral (bottom row) projection for the five CT image pairs. The base of each vector represents the position of a landmark feature in the maximum inhale phase from each 4D CT, while the head represents the corresponding feature location in the respective maximum exhale phase. 1280, 1487, 1166, 1561, and 1268 individual landmarks were manually selected for cases 1 through 5, respectively. Color scales are provided for each case indicating magnitude displacements in units of millimeters.

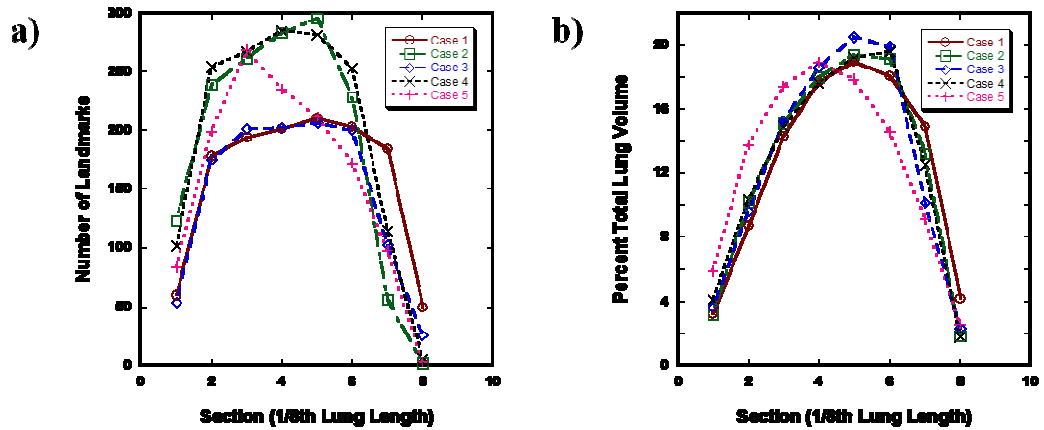
The stated goal of >10 feature pairs per lung per axial image slice served as a guideline to ensure uniform spatial distribution of the validation landmarks. In practice, the number of landmarks required to adequately sample a given image slice may be more or less, depending on the volume of lung contained within that slice. To demonstrate this effect, the inhale lung voxels for each case were segmented based on simple histogram segmentation and three-dimensional connectivity. The lung volumes were then partitioned into blocks, each approximately  $1/8^{\text{th}}$  of the total number of axial slices containing lung.

|                       | Case 1      | Case 2      | Case 3      | Case 4      | Case 5      |
|-----------------------|-------------|-------------|-------------|-------------|-------------|
| # Right lung          | 672         | 767         | 637         | 803         | 661         |
| # Left lung           | 608         | 720         | 529         | 758         | 607         |
| # Total               | 1280        | 1487        | 1166        | 1561        | 1268        |
| Avg Displacement (mm) | 4.01 (2.91) | 4.65 (4.09) | 9.42 (4.81) | 6.73 (4.21) | 7.10 (5.14) |
| Max Displacement (mm) | 12.65       | 17.8        | 21          | 18.46       | 24.78       |

**Table 2.1. Reference landmark characteristics.** The number of expert-determined landmark feature points is shown for each case in terms of right, left and total lung points. Average (standard

deviation) and maximum landmark displacements are seen to vary substantially across the five datasets.

Figure 2.3a shows the number of landmarks contained within each of the sectioned superior-inferior blocks for all cases, while figure 2.3b shows the corresponding distribution of total lung volume over the same superior-inferior extent. Though the volume measurements are only approximations based on image segmentation, figure 2.3 suggests that the quantity of selected feature points as a function of location in the superior-inferior direction is primarily attributable to the superior-inferior distribution of lung volume. Note that figure 2.3b does not provide any indication as to the relative lung volumes between cases.



**Figure 2.3. Superior-inferior distribution of landmark points and lung volume.** Each lung volume was partitioned into blocks approximately  $1/8^{\text{th}}$  the total lung length (i.e., total axial slices). **a)** The cumulative number of landmarks is shown for each case as a function of the sectioned superior-inferior extent of the lung. **b)** The percentage of total lung volume within each section is shown for all cases.

#### 2.2.4 Landmark Selection Variability

In order to provide estimates of reproducibility of target point selections, random samples of 200 source feature points were generated for each case from the primary landmark sets. The source lists were then imported into the APRIL interface and re-registered by two secondary readers to estimate inter-observer reproducibility. The primary reader also re-registered the sampled sources to estimate intra-observer reproducibility. Each of the

repeated registrations was performed independently and without prior knowledge of the primary target point selections.

The repeated registration error was quantified as the three-dimensional Euclidean distance between the original target point in the primary data set and the corresponding point selected in the repeated registration. Repeated registration errors by the secondary readers were combined to estimate inter-observer reproducibility, while the intra-observer estimates were determined from repeated registrations by the primary reader. Mean errors and corresponding standard errors were calculated for each case, as well as over the combined set of error measurements. The observed error distributions across all cases for both the primary and secondary readers were skewed with respective skewness parameter values of 1.3 and 3.2. Two-sample *t*-tests were performed comparing mean inter- and intra-observer repeated registration errors. Although the error distributions were skewed, the Central Limit Theorem [64] justifies a *t*-test for comparison of means. A summary of the error measurements is provided in table 2.2.

|                  | <b>Case1</b>      | <b>Case2</b>      | <b>Case3</b>      | <b>Case4</b>      | <b>Case5</b>      | <b>Combined</b>   |
|------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| <b>Primary</b>   | <b>0.88(0.07)</b> | <b>0.61(0.07)</b> | <b>1.11(0.07)</b> | <b>0.73(0.07)</b> | <b>0.84(0.08)</b> | <b>0.83(0.03)</b> |
| <b>Secondary</b> | <b>0.83(0.07)</b> | <b>0.74(0.05)</b> | <b>1.14(0.07)</b> | <b>0.79(0.05)</b> | <b>0.95(0.06)</b> | <b>0.89(0.03)</b> |
| <b><i>p</i></b>  | <b>0.6473</b>     | <b>0.1562</b>     | <b>0.7617</b>     | <b>0.4758</b>     | <b>0.2691</b>     | <b>0.2079</b>     |

**Table 2.2. Landmark reproducibility summary.** Inter- and intra-observer reproducibility of individual target point selections were estimated from repeated registration of uniform sampled of 200 source feature points for each case. Mean (standard error) errors were also determined for the set of repeated registrations. Two-sample *t*-tests were performed comparing mean inter- and intra-observer repeated registration errors (with *p*-values shown). All error measurements are presented in units of millimeters.

In four out of the five cases, mean intra-observer registration error was lower than mean inter-observer error, though the differences did not reach statistical significance for any of the cases ( $p \geq 0.1562$ ). Mean repeated registration errors (SE) by the primary reader ranged from 0.61 (0.07) – 1.11 (0.07) mm for the five cases, while mean inter-observer errors ranged from 0.74 (0.05) – 1.14 (0.07). Over the combined set of 3000 repeated registration measurements, mean error for all observers was 0.87 (0.02) mm, with inter-quartile range 1.16 mm.

### 2.2.5 *Spatial Localization Uncertainty*

Finite sampling and image resolution due to acquisition and reconstruction inherently impose fundamental uncertainties associated with spatial localization of anatomical landmarks in medical images. For a fixed image resolution and voxel size, an observer cannot meaningfully localize a prominent feature point with sub-voxel accuracy. Thus, the manually determined landmark correspondences are described by integer coordinate pairs. In general, when quantifying landmark-based registration errors one must take into account that there is an inherent spatial uncertainty associated with the voxel localization of each landmark feature that is a function of the voxel dimension in each direction. For the CT images utilized in this study, the maximum RL and AP voxel dimensions were 1.16 mm. The SI voxel dimension for each case was 2.5 mm. However, because of the large number of landmarks included in our analysis, we were able to estimate the average error associated with the landmark identification by both the readers and the registration algorithms with sub-voxel accuracy. To understand why this is possible, recall that a Bernoulli proportion can be estimated with arbitrary precision (for a sufficiently large sample size), even though the outcome of each Bernoulli trial takes on only one of two discrete values (0 and 1). Furthermore, we require as a specific criterion for the manual selection of point pairs, that the image features are identifiable in both source and target images. In this sense, a secondary effect of image resolution on landmark selection is on the quantity of feature points satisfying this criteria for a given image pair. Thus, for relatively poor resolution images fewer usable landmarks can be identified.

## 2.3 **Landmark-based Evaluation of Deformable Image Registration**

The goal of deformable image registration is to find a point-to-point correspondence between two given images. This desired correspondence should relate the location of each underlying tissue element represented in each voxel in the first image to that in the second image. In order to demonstrate the utility of the large landmark sets as a means for assessing DIR performance, two deformable image registration algorithms were implemented, providing example DIR output for the five patient cases described above. The two methods are briefly described.

### 2.3.1 *Optical Flow*

Optical flow methods (OFM) [65] comprise a large class of image registration techniques where the voxel correspondence is determined by computing a velocity field describing

the apparent motion depicted in the two images. For a single pair of images, the velocity field is equal to the displacement field with the time step assumed to be unity. Several reviews of these methods exist, as do studies that focus on the performance of different optical flow implementations and techniques (see [66], for example). In a previous work, we employed optical flow to track tumor motion and calculate ventilation from 4D CT [15, 20]. Our optical flow implementation is based on an iterative procedure [65] used to solve for the unknown velocity at each voxel:

$$v_{n+1} = \bar{v}_n + \nabla I \left( \frac{\nabla I \cdot \bar{v}_n + \frac{\partial I}{\partial t}}{\alpha^2 + \|\nabla I\|^2} \right), \quad (3)$$

where  $n$  and  $n+1$  are iteration counts and  $\bar{v}_n$  is the average velocity taken over the nearest neighboring voxels. This method is equivalent to the well-known Gauss-Seidel method [67] where the latest available velocity values are used in calculating the average. All necessary temporal and spatial image derivatives are approximated with finite differences applied to the two given images. In this study, eight iterations of equation (3), with  $\alpha = 25$  were performed for all OFM registrations. To ensure variability in the DIR spatial accuracy performance, no attempts were made to optimize individual case registrations.] "

### 2.3.2 Landmark-based Deformable Image Registration

Landmark-based algorithms represent an alternative class of image registration techniques in which sets of registered control point pairs are used to calculate an interpolating function that estimates the displacement of all voxels within the volume of interest (VOI). The point correspondence process is divided into

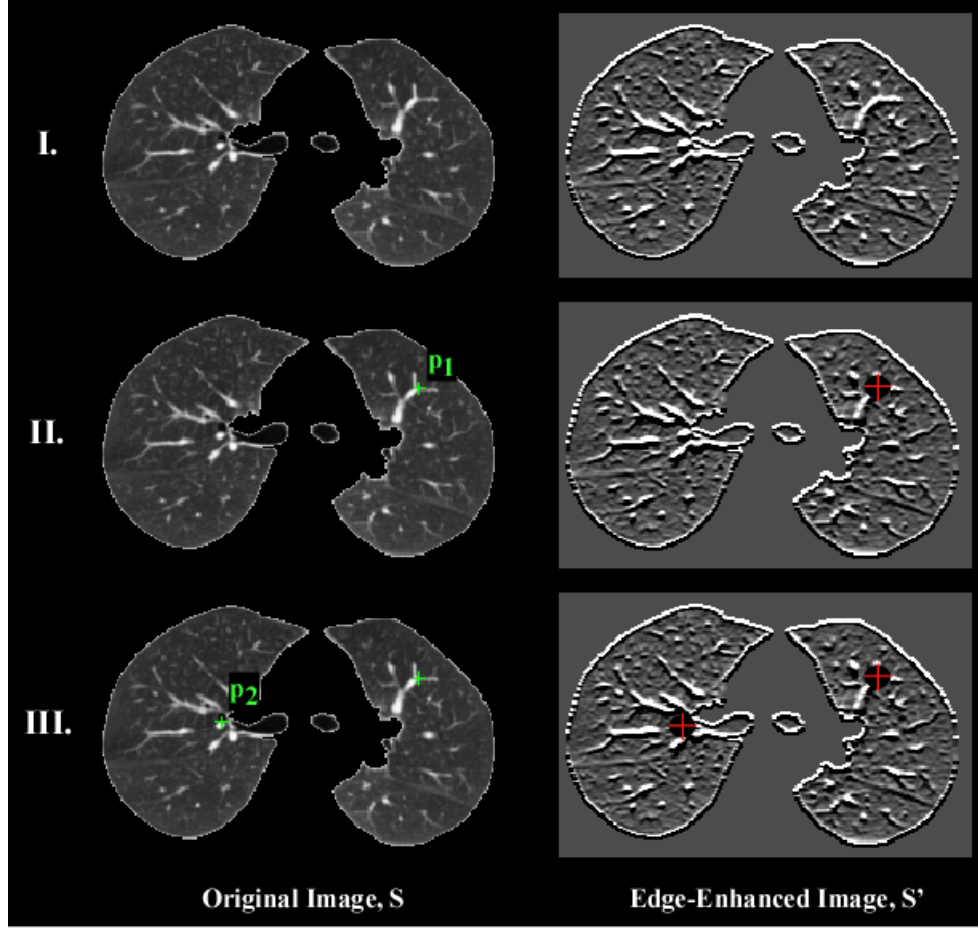
1. automated segmentation of the lungs to form the lung VOI,
2. identification of high contrast source feature control points, and
3. local neighborhood correlation to find the corresponding target volume points.

A pulmonary segmentation algorithm [68] based on histogram segmentation and connectivity was applied to each CT image volume. The resulting lung parenchyma VOI was used in subsequent analysis. An edge enhanced image generated using the three-dimensional Sobel operator is used to identify feature points within the source image  $S$  that lie on image discontinuities. Source

control points  $(p_i)$ , distributed over the lung surface and interior volume, are then selected automatically according to the algorithm illustrated in figure 2.4.

Given the combination of interior and perimeter landmark control points, an initial set of corresponding target landmarks  $(q_i)$  is determined based on a weighted cross correlation of the local source feature neighborhood with a larger search region in the target volume. Consider the source landmark  $p$  located at position  $(p_x, p_y, p_z)$  in the source image. Isotropic local neighborhoods  $N_s$  and  $N_T$ , symmetric about  $p$  are generated in the respective source and target images, with  $N_T \gg N_s$ . Let the width of  $N_s$  be given by  $w_N = 2\eta + 1$ , where  $\eta \in \mathbb{Z}$  is a dummy variable introduced for notational simplicity. Then for all  $(x, y, z) \in N_T$  we calculate the cross correlation coefficient  $C$  given by:

$$C(x, y, z) = \sum_{i=-\eta}^{\eta} \sum_{j=-\eta}^{\eta} \sum_{k=-\eta}^{\eta} N_T(x+i, y+j, z+k) N_s(p_x+i, p_y+j, p_z+k). \quad (4)$$



**Figure 2.4. Volume-based control point selection.** A two-dimensional example is shown in transaxial section illustrating the automated source control point selection algorithm. **I)** An edge-enhanced image  $S'$  is first created from the original image  $S$ , by  $S' = S * \Omega$ , where  $*$  is convolution and  $\Omega$  is the Sobel operator. **II)** The location of the maximum intensity pixel in  $S'$ , denoted by the red cross, is initially selected as the first source feature  $p_1$ . A spherical region of radius  $r$  centered on  $p_1$  in  $S'$  is set to zero to control spacing of the feature points. **III)** The maximum intensity pixel in the modified  $S'$  is selected as the second source feature point  $p_2$ , and a second spherical region in  $S'$  similarly set to zero. The process is repeated until the maximum intensity in the modified  $S'$  reaches a predefined user-determined threshold. The black-out radius  $r$  is set to vary automatically and depends on the desired quantity of source feature points requested by the user.

Ideally, the maximum value of  $C$  would represent the best match for  $N_s$  within the target search region. However, since there is not a guaranteed unique global maximum for  $C$ , a new function

$\hat{C}$  is created to penalize candidate target voxels as a function of magnitude distance from the given source position:

$$\hat{C}(x, y, z) = C(x, y, z) - \alpha \|(x, y, z) - (p_x, p_y, p_z)\| \quad (5)$$

where  $\alpha$  is a weighting factor. The best estimate of the corresponding feature position in the target image is then given according to the maximum value of  $\hat{C}$ . This process is repeated for all  $p_i$  in the set of source landmark features. In this study, all automated point pairs were visually inspected and manually adjusted as necessary prior to proceeding with the volumetric image registration. Finally, given the set of feature points  $p_i$  in the inhale (source) image and corresponding  $q_i$  in the exhale (target) image, the last task is to construct a smooth deformation  $f(\vec{x})$ , for all lung voxels that approximates or interpolates the input point pairs;  $\|f(p_i) - q_i\| < \varepsilon$ , where  $\varepsilon > 0$ , or  $f(p_i) = q_i$ . To construct the final deformation, we utilize a purely geometric method based on least squares that is simple to implement and very fast to evaluate. This method, known as moving least squares (MLS)[69], constructs the desired  $f(\vec{x})$  as the solution to a weighted least squares problem. To compute  $f(\vec{x})$  at an arbitrary position  $v$ , we solve for an affine function  $A_v(x)$  that minimizes the expression:

$$\min \sum_i w_i \|A_v(p_i) - q_i\|^2, \quad (6)$$

where  $p_i$  and  $q_i$  are the  $i^{th}$  pair of corresponding source and target feature positions, and the weights  $w_i$  are of the form:

$$w_i = \frac{1}{\|p_i - v\|^2 + \varepsilon}, \quad (7)$$

and  $\varepsilon \geq 0$ . Using this approach, the spatial transformation function can be expressed simply as:

$$f(\vec{v}) = A_v(\vec{v}). \quad (8)$$



That is, for each voxel  $v$  in the source image,  $f(\vec{v})$  is a different transformation, dependent on the set of weights  $w_i$ . Note that for  $\varepsilon = 0$ , the function  $f_v$  interpolates  $q_i$  as  $v$  approaches  $p_i$ .

### 2.3.3 Evaluation of Deformable Image Registration

" [Fundamentally, spatial registration error is defined as the difference between a calculated output and the designated reference standard. In this case, large sets of manually delineated feature pairs serve as the primary validation data. For this comparison to be strictly valid, the evaluation of manual and calculated landmark registration should be equivalent. That is, since an observer selects integer voxel locations in an image pair as corresponding point sets, the comparison with calculated positions should also be performed on the same integer grid. This is achieved simply by rounding the final displaced position of each coordinate of interest to the nearest integer. As described in section 2.2.5, we were able to estimate the average error associated with the landmark identification by the registration algorithms with sub-voxel accuracy, due to the large measurement sample sizes. Numerically, the mean errors determined from the rounded and floating point DIR positions will likely be similar. This is due to the fact that on average approximately equal quantities of test voxels are rounded toward their respective reference target position as are rounded away. However, to ensure equivalence of the reference standard and the calculated outputs, integer positions should be utilized.

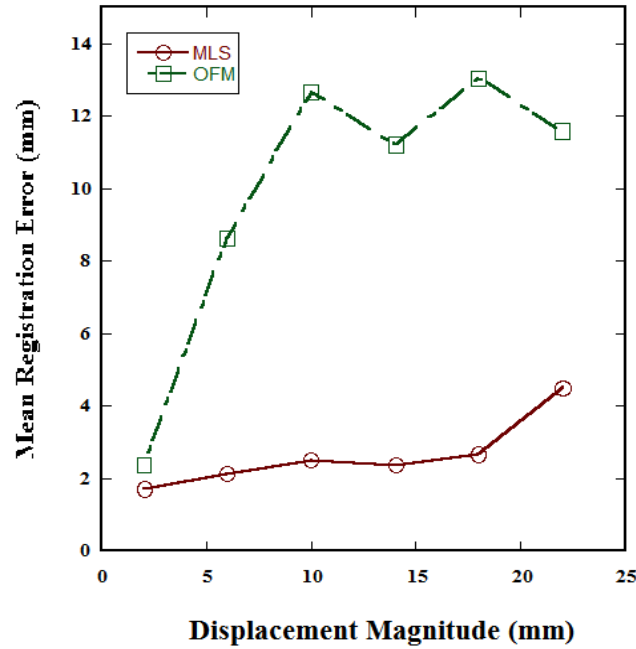
Point registration error was quantified as the three-dimensional Euclidean distance between target voxels in the primary data set, and those determined by applying the calculated DIR transformation to the corresponding source feature location. Mean registration error and corresponding standard error were determined for both DIR algorithms over the set of validation landmarks, providing a global measure of spatial accuracy performance for each case. Mean errors were also determined over the combined set of expert-determined feature points for all cases. Additionally, errors were assessed separately for individual RL, AP, and SI component directions. Two-sample  $t$ -tests were performed to assess the statistical significance of differences in mean registration error between algorithms. Skewness parameter [64] values for OFM and MLS were 1.9 and 2.3, respectively. As above, the  $t$ -test was justified for comparison of the mean errors by the Central Limit Theorem [64]. For both methods, since the observed error distributions were skewed, the non-parametric Spearman rank correlation

coefficient was calculated to quantify the statistical correlations between registration error and each of displacement magnitude, change in intensity between image pairs, and local contrast within a  $(5 \times 5 \times 5)$  voxel neighborhood surrounding each source feature point.

Table 2.3 summarizes the spatial accuracy performance of the two DIR algorithms for registration of the five thoracic CT image pairs. Over the complete validation landmark set, mean registration errors (SE) for respective OFM and MLS DIRs were 6.90 (0.10) and 2.05 (0.02) mm. Inter-quartile ranges for the OFM and MLS DIR were 10.03 and 1.63 mm, respectively. For the OFM DIR, mean RL, AP, and SI component errors were each greater than 2 mm, with the largest registration errors occurring in the AP direction. In contrast, all MLS mean component errors were less than 2 mm, with the largest occurring in the SI direction. Spearman's rank correlation coefficients were calculated to assess correlation of registration error with landmark displacement magnitude, change in intensity, and local source landmark contrast. For both algorithms, the largest correlation was observed for displacement magnitude, with corresponding OFM and MLS Spearman coefficients of 0.562 and 0.203, respectively.

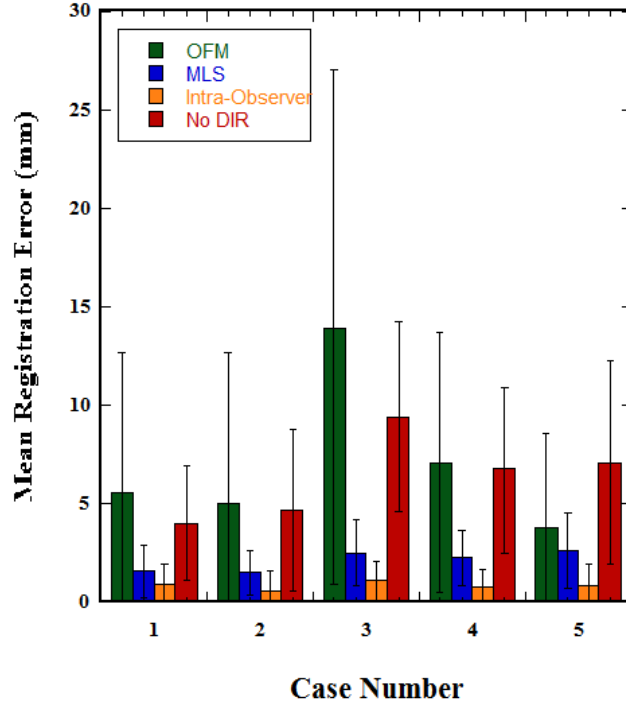
|                                    | Optical Flow | Moving Least Squares | <i>p</i> -value |
|------------------------------------|--------------|----------------------|-----------------|
| <i>Mean Error (standard error)</i> |              |                      |                 |
| 3D                                 | 6.90 (0.10)  | 2.05 (0.02)          | < 0.001         |
| Right-Left                         | 2.21 (0.04)  | 0.77 (0.01)          | < 0.001         |
| Anterior-Posterior                 | 4.94 (0.09)  | 0.90 (0.01)          | < 0.001         |
| Superior-Inferior                  | 2.78 (0.05)  | 1.05 (0.02)          | < 0.001         |
| <i>Spearman Rank Correlation</i>   |              |                      |                 |
| Magnitude Displacement             | 0.562        | 0.203                | < 0.001         |
| Local Source Contrast              | -0.100       | -0.12                | < 0.001         |
| Intensity Change                   | 0.107        | 0.014                | 0.2578          |

**Table 2.3. DIR spatial accuracy comparison.** Three dimensional and component mean registration errors derived from the complete validation point set are shown for both DIR algorithms. Two-sample *t*-tests were performed to assess the difference in mean errors, with corresponding *p*-values shown. Spearman's rank correlation coefficient was also determined to investigate correlation of the measured spatial error with each of magnitude landmark displacement, local source feature contrast, and change in reference landmark intensity between images. All registration errors are shown in units of millimeters.



**Figure 2.5. Registration error versus displacement magnitude.** For each DIR algorithm, registration errors were binned corresponding to magnitude displacement of the reference landmarks in 4 mm increments. Mean registration errors were then determined for each bin and plotted versus displacement magnitude.

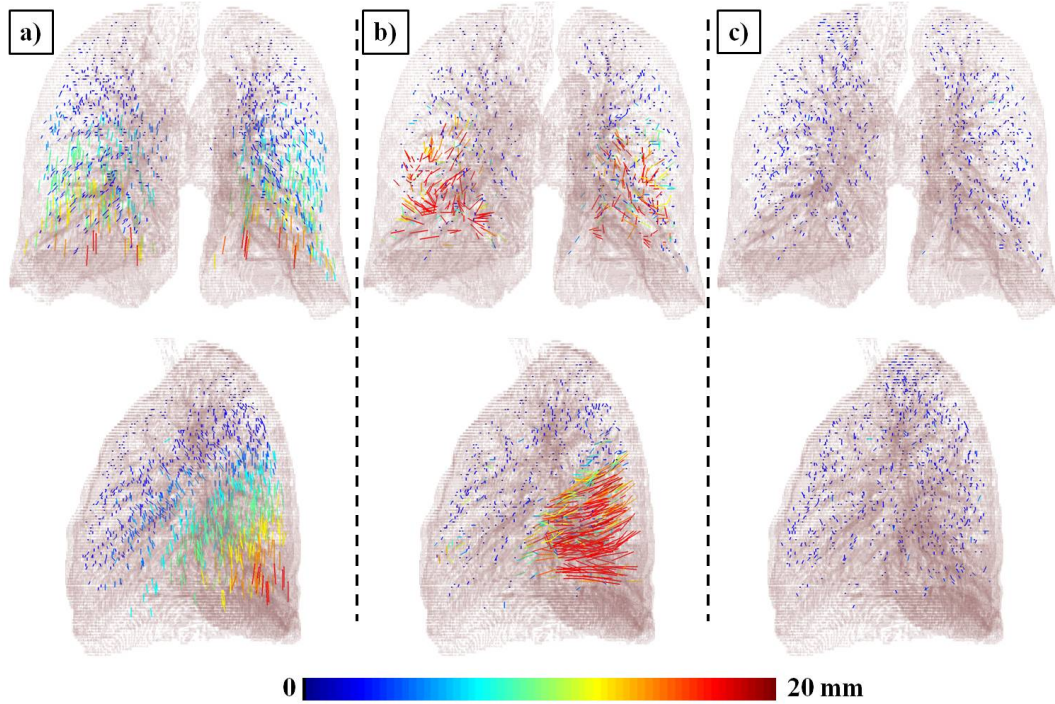
The summary data presented in table 2.3 is useful for providing a global assessment of the spatial accuracy characteristics of the two DIR algorithms. However, more detailed investigation of the nature of the errors can easily be performed through graphical interpretation of the same results by binning the error measurements appropriately. For example, the correspondence between registration error and landmark displacement is depicted graphically for both algorithms in figure 2.5. Registration errors were binned according to magnitude displacement of the complete set of 6762 validation landmarks in 4 mm increments. The figure shows relatively consistent behavior of the MLS registration errors over the range of displacement magnitudes, while a positive trend clearly can be seen for the OFM errors, consistent with the calculated Spearman correlation coefficients.



**Figure 2.6. Case registration errors.** Mean OFM and MLS DIR registration errors ( $\pm 1$  standard deviation) are shown for each case. Corresponding intra-observer repeated registration errors, as well as landmark displacement (i.e., no DIR) are also shown for reference. Those cases for which DIR registration error exceeds the corresponding landmark displacement magnitude indicate instances in which deformable registration of the image pairs resulted in increased misalignment of the validation landmarks.

Figure 2.6 shows a bar graph depiction of the mean registration errors for each individual case. The corresponding intra-observer repeated registration errors, as well as the mean landmark displacement magnitudes are similarly shown for reference. Mean case errors ranged from 1.47 (0.03) – 2.55 (0.05) mm for MLS DIR and 3.73 (0.14) – 13.96 (0.38) mm for the OFM DIR. Those cases for which DIR registration error is greater than the corresponding landmark displacement magnitude indicate instances in which deformable registration of the image pairs resulted in increased misalignment of the landmarks. It is important to note that there is no indication of this in the summary error statistics presented in table 2.3. Numerical spatial accuracy measurements can only be properly interpreted with reference to the validation data from which the measurements were acquired. Furthermore, reference should also be made to the inter- and intra-observer variance obtained during characterization of the landmark datasets. An

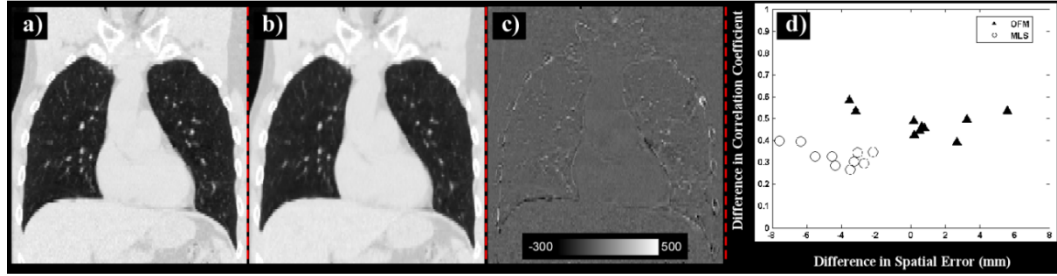
algorithm that achieves a statistically indistinguishable result when compared to the expert landmarks effectively reaches the maximum resolution of the dataset.



**Figure 2.7. Residual error renderings.** a) 1487 pulmonary landmark features were manually registered between the maximum inhale and exhale component phase volumes from a 4D CT (case #2). The expert-determined displacement vectors are shown projected onto a surface rendering of the inhale lung volume. Residual error vectors are shown for b) OFM and c) MLS algorithms. Each error vector is shown pointing from the manually delineated feature location in the target image to that determined from the respective DIR transformation.

Volume renderings of the lung surfaces were also generated and overlain with a vector representation of each of the individual error measurements to visually assess the spatial distribution of registration errors within the anatomic context. Figure 2.7a shows the vector representation of the expert-determined validation landmark point set for an example case, projected onto a surface rendering of the corresponding inhale CT lung voxels. The residual error vectors for both DIR algorithms are shown in figure 2.7b&c. In these figures, error vectors are shown pointing from the manually delineated feature location in the target image to that determined from the respective DIR transformation. The OFM plot shows a relatively large AP component error, consistent with the global

assessment presented in table 2.3. Graphically, no systematic tendencies are apparent for the MLS DIR, suggesting little correlation of registration error with spatial location. More detailed graphical or quantitative error analyses within specific regions of the lung volume, for example on an individual lung lobe basis, can be achieved simply by applying binary masks of the desired ROIs to the raw error measurement data.



**Figure 2.8. DIR performance metrics.** Coronal CT slices are shown for an example case from **a)** the original CT data next to **b)** the corresponding slice from the estimated inhale image derived from the optical flow DIR. **c)** The difference image is also shown. Visual and quantitative assessment of image similarity following DIR can result in potentially misleading evaluation of DIR spatial accuracy performance. For the volumetric image pair depicted, mean registration error was 4.98 (SD: 7.66, Max: 41.76) mm. **d)** Difference in correlation coefficient is shown versus corresponding difference in landmark registration error before and after DIR. Lung voxel ROIs were determined from the set of inhale images, separately for individual right and left lungs in order to increase measurement sample size. Positive change in correlation coefficient indicates increased image similarity following DIR, while negative change in spatial error indicates improved alignment of validation landmarks. Note that increased correlation coefficient does not necessarily imply improved spatial accuracy.

Finally, a standard image intensity-based measure of DIR performance [21, 22, 25, 37, 38] was also determined for comparison with the spatial accuracy measurements derived from the validation point sets. For each case, estimated inhale image volumes were generated by applying the calculated DIR transforms to the inhale voxel grid, and performing tri-linear interpolation of the mapped exhale neighborhood intensities to determine the estimated intensity of each voxel. A coronal slice from an example inhale image is shown in figure 2.8a, next to the corresponding slice from the estimated inhale image derived from the optical flow DIR (figure 2.8b). The difference image is also shown in figure 2.8c. Visually, the images appear similar. However, visual inspection

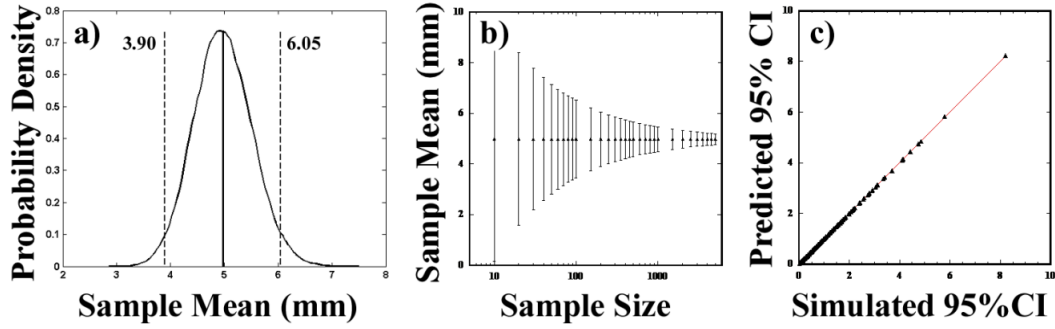
alone provides no indication of the underlying DIR spatial accuracy. Mean registration error over the set of validation landmarks for the case depicted was 4.98 (SD: 7.66, Max: 41.76) mm.

Correlation coefficients were then calculated to assess the quantitative gray-scale similarity of the inhale and estimated inhale volumes. To avoid the influence of background voxel intensities and to determine correlation coefficients on an individual lung basis, the calculation was masked by separate right and left lung regions of interest (ROIs) determined from the original inhale image. Using the same lung ROIs, correlation coefficients were similarly calculated between the inhale and unregistered exhale volumes to determine the effect of DIR on image similarity. In figure 2.8d, the calculated changes in correlation coefficient are shown plotted versus the difference in mean registration error of the validation landmarks before and after DIR to graphically assess correspondence of the two performance metrics. In this example, positive change in correlation coefficient indicates an increase in image similarity within the lung ROI following DIR, while positive change in spatial error indicates an *increase* in misalignment of the validation landmarks. The two DIR algorithms exhibit different behavior; a greater increase in correlation was found with the OFM algorithm and a consistent reduction in spatial error was found with the MLS algorithm. For a majority of the test cases, the OFM algorithm resulted in an increase in the spatial error. The lack of correlation between image similarity and DIR spatial error is a new finding. For both algorithms, DIR consistently resulted in improved image similarity within the lung ROIs. However, the increase in correlation provided no indication of the underlying spatial error, which was made worse in some cases. For the data presented in this study, the correlation coefficient fails to provide even a reliable measure for relative performance between algorithms. For objective evaluation of the spatial accuracy of the calculated displacement of individual volume elements, intensity-based metrics afford little useful insight, as no information is provided regarding the origin of the aligned voxel intensities (regardless of their equality). Hence, correlation and gray-scale similarity measures [21, 22, 25, 37, 38] and/or visual checks [70] are inadequate for evaluation of DIR results.

## **2.4 Landmark Sample Size Analysis**

Objective evaluation of DIR based on large samples of landmark point sets can be a highly effective and informative strategy for characterization and comparative evaluation of algorithm performance. However, the large landmark datasets represent more than a

useful tool when detailed assessment of DIR is desired. Rather, they should be considered a statistical necessity when the spatial accuracy characteristics of a given algorithm are not established *a priori*. In this section, we utilize the statistical properties of the two sets of DIR outputs over the validation point sets to demonstrate the effect of landmark sample size on the uncertainty associated with spatial error estimation.

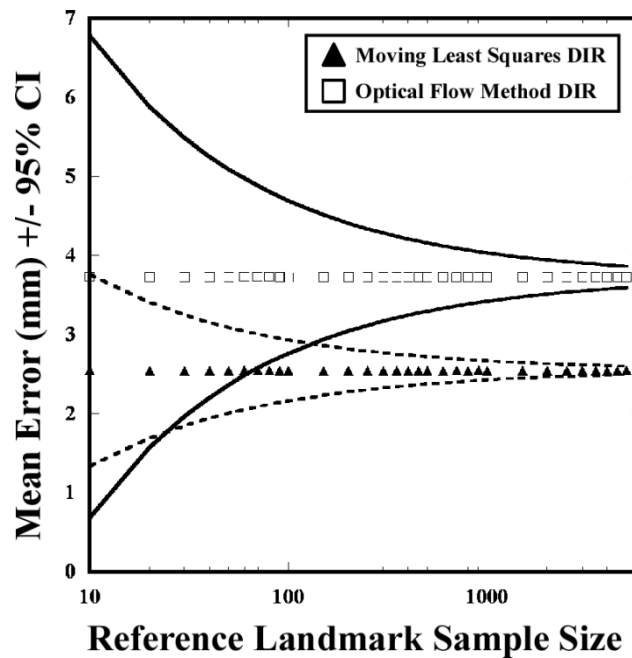


**Figure 2.9. Mean registration error uncertainty.** **a)** The experimentally determined distribution of 100,000 mean registration errors is shown for an example case (case #5), for fixed validation sample size of 200. Vertical bars indicate sample mean  $\pm$  95% CIs. **b)** The distribution illustrated in a) was similarly determined for validation sample sizes ranging from 10 to 5,000. Corresponding sample means  $\pm$  95% CIs are shown in a semi-log plot, demonstrating the effect of sample size on the statistical uncertainty associated with mean registration error. **c)** For all cases and both DIR algorithms, simulated versus predicted values for 95% CIs were plotted over the range of experimental sample sizes to assess linear correlation, with corresponding  $R^2 = 0.99$ .

For both algorithms, cumulative distribution functions (CDFs) were generated from the corresponding set of error measurements for each case. To simulate the spatial error information derived from validation point sets of different size, uniform samples of the individual CDFs were obtained for sample sizes ranging from 10 to 5,000. For each sample size, 100,000 independent sample sets were obtained. At each sample size increment, an independent calculation of the mean spatial error was performed for each of the 100,000 error samples. The distribution of sample means was then used to determine the expected mean spatial error  $\pm$  95% confidence intervals (CIs). Figure 2.9a shows the distribution of the 100,000 experimentally determined mean registration errors for the fixed validation sample size of 200 for an example case. The corresponding 95% CIs are also indicated. Figure 2.9b shows the corresponding measurements for sample sizes ranging from 10 to 5,000. The set of CIs obtained through simulation was then



compared to predicted intervals derived from basic statistics considerations. A thorough description of the statistical framework for uncertainty estimation can be found in most introductory textbooks in statistics or data analysis (for example, [71]). For each case and for both DIR algorithms, respective values for the standard deviation of error measurements were utilized to evaluate predicted values for the 95% CIs on the mean registration error for the experimental range of sample sizes described above. The combined set of predicted versus measured values for 95% CIs is plotted in figure 2.9c. The square of the Pearson correlation coefficient,  $R$ , was calculated to assess linear correlation of the simulated and predicted CIs, with  $R^2 = 0.99$ .



**Figure 2.10. Comparative evaluation uncertainty.** A semi-log plot of predicted uncertainty in mean registration error versus sample size is shown for an example case. Predicted values for 95% CIs were determined for validation sample sizes ranging from 10 to 5,000. For both DIR algorithms, the 95% CIs on the mean registration error are shown. For this case, a sample size of approximately 150 validation landmarks is required to obtain non-overlapping 95% CIs. Comparative evaluation based on fewer than the required landmarks increases the probability that the comparison is a misrepresentation of the relative DIR spatial accuracy performance.

In practice, the statistical uncertainties associated with mean registration error depicted in figure 2.9 can lead to potentially misleading assessment of DIR spatial

accuracy characteristics. Figure 2.10 illustrates this point in the context of comparative evaluation of DIR spatial accuracy between two algorithms. For an example case, mean  $\pm$  predicted 95% CIs are shown as a function of sample size for both OFM and MLS DIR. For this example, a minimum of approximately 150 uniformly distributed validation landmarks are required to obtain non-overlapping 95% CIs. Comparison of mean registration errors based on fewer than the required landmarks increases the probability that the comparative evaluation is a misrepresentation of the relative spatial accuracy characteristics of the two algorithms.

In general, the sample size required to obtain non-overlapping CIs will vary across test cases. Thus, for a given algorithm, it would be beneficial to formulate an estimate of the sample size required to obtain 95% CIs of a specified length (e.g. 1 mm) on the mean registration error. To do so requires incorporating all available error information pertaining to a given algorithm in order to calculate a pooled standard deviation of error measurements obtained from all available cases. The pooled standard deviation,  $\bar{s}_{DIR}^p$ , is given by:

$$\bar{s}_{DIR}^p = \left[ \left[ \sum_{i=1}^C (N_i - 1) \right]^{-1} \left[ \sum_{i=1}^C (N_i - 1) s_{i,DIR}^2 \right] \right]^{0.5}, \quad (9)$$

where  $C$  is the total number of available cases,  $N_i$  is the validation sample size for the  $i^{th}$  dataset, and  $s_{i,DIR}$  is the corresponding standard deviation of error measurements associated with algorithm ‘DIR’. If we require a 95% CI within a specified range ( $\pm d$  mm) of the mean error, then the necessary reference sample size is approximated by:

$$N_{\pm d}^{DIR} = \left( \frac{2\bar{s}_{DIR}^p}{d} \right)^2. \quad (10)$$

This is an important point to consider in the interpretation of DIR spatial accuracies reported in the literature. The resulting sample size is not necessarily a large number (for example, for  $\bar{s}_{DIR}^p = 2$  mm and  $d = 1$  mm,  $N^{DIR} = 16$ ). Rather, it represents a minimum statistical requirement to ensure the mean DIR registration error is an accurate representation of the spatial accuracy performance for any given case. Furthermore,  $\bar{s}_{DIR}^p$

is generally not known *a priori*, and therefore itself must be estimated from large samples of uniformly distributed measurements. It is highly recommended that data sets that are made publicly available for purposes of multi-institutional comparative evaluation studies consist of sufficiently large validation test points to avoid drawing erroneous conclusions based on insufficient data. In the design of such studies, estimates of respective values for  $\bar{s}_{DIR}^p$  should be obtained for all  $n$  participating algorithms from prior evaluation studies, incorporating associated uncertainties, to calculate appropriate validation landmark sample sizes. For the specified interval range,  $d$ , the minimum allowable sample size used for the specific study should then be given by  $\max \{N_{\pm d}^{DIR_1}, \dots, N_{\pm d}^{DIR_n}\}$ . We propose the  $\bar{s}_{DIR}^p$ , a characteristic of the given algorithm and the anatomic target, be measured on acceptance testing or comparative evaluation of new DIR algorithms. It is also worthy to note, that the same sample size considerations apply regardless of whether the validation test points are delineated in patient or phantom images.

## 2.5 Discussion

Expert determined landmark correspondences have become a widely adopted reference for evaluating DIR accuracy for lung image data, however there has been great variability in their use. In this study we have presented a framework for objective evaluation of thoracic deformable image registration spatial accuracy, based on the use of large samples of expert-determined landmark feature pairs between volumetric images as a reference for spatial accuracy measurements. A summary of the methodology is presented:

### 2.5.1 Selection of Anatomical Landmark Pairs

The use of registered landmarks as an objective metric for evaluation of image registration loses its significance if the point correspondences are calculated automatically. Thus, it is crucial that the individual feature points are first selected then manually registered between image volumes by a human observer, with expertise in imaging of the appropriate anatomic site. This is undoubtedly a lengthy task, and it is difficult to appreciate the necessity of enduring the process, without having some prior demonstration as to why there is a necessity. Thus, there has been no compelling reason for investigators to pursue what would presently be considered as unnecessarily large data sets. In fact, large data sets are crucial, thus the process should be streamlined as best

as possible, while still leaving the actual registration of individual feature points in the hands of the expert. Simple software design considerations can be highly effective in this regard.

The APRIL software utilized here was developed to maximize the number of landmark point pairs that a user can select. For any designated source landmark, optional computer assistance tools provided rapid localization of an estimated target correspondence, based on user-determined threshold and search range criteria. However, the final selection of the corresponding point was performed manually by the user to ensure the selection represents the expert choice and not the particular calculated estimate. A range of 1166 to 1561 unique anatomical features were manually identified and tracked between the five individual pairs of treatment planning CT images.

For summary statistics such as mean and standard deviation of the measured registration errors to accurately reflect the DIR performance throughout the lung, points must be distributed sufficiently uniformly in space, such that spatial variations in the DIR accuracy are detected. Perhaps the most significant perceived drawback regarding the use of manually registered feature points for objective evaluation of DIR is the notion that naturally occurring anatomical features are too few and too unevenly distributed to provide for rigorous performance evaluation. One reason for this view may be the requirement by some investigators that the anatomic identification of each landmark point is necessary (e.g. the  $n^{\text{th}}$  generation of the right main bronchus), similar to landmark registration in neuro-imaging applications. In contrast, we feel the expert user must simply uniquely identify corresponding image features without identifying their exact anatomic location.

### *2.5.2 Characterization of Landmark Datasets*

Estimates of variability within the primary reader (intra-observer) and among readers (inter-observer) for matching the corresponding landmark features must be obtained. In general, the variance sets a lower limit on the spatial accuracy that is detectable using the validation landmarks. This characterization process also requires large samples of measurements to ensure tight confidence intervals on the estimates of observer variance.

The complete reference displacement set consisting of five lung CT image pairs and 6762 landmark point pairs was statistically characterized with measurements of the intra- and inter-observer variance by repeated registration of multiple subsets of feature points. An important factor not specifically investigated in this study is the effect of an

observer's experience or familiarity with the APRIL software on the manual registration process. This is an important point to consider because the ability to resolve registration errors is largely a function of how well the validation points can be reproduced. In practice, care should be taken to ensure adequate training in the manual registration process prior to the acquisition of formal repeated registration measurements. A more thorough characterization of the observer variance based on larger populations of participating observers is still necessary.

### *2.5.3 Evaluation of Deformable Image Registration*

The sets of validation landmarks were utilized to perform quantitative comparative evaluation of a gradient-based OFM algorithm and a landmark interpolation algorithm based on MLS [69]. The validation landmarks provided for statistical tests on mean registration errors, as well as visual and quantitative assessment of spatial accuracy performance with location and magnitude displacement. It should be emphasized that the goal of this study was not to perform an explicit comparison between landmark-based MLS and gradient-based OFM for thoracic DIR. Though the OFM results presented here were indeed poor, further improvement based on optimization of internal parameters such as the regularization smoothing parameter ( $\alpha$ ) and the iterations of equation(3), could almost certainly be achieved. Furthermore, the spatial accuracy of the MLS DIR is a function of the quantity and uniformity of the input point pairs used for MLS interpolation. Thus, variations in input landmark selection will result in variable DIR output. Optimization of input parameters for both algorithms could be investigated further based on the evaluation methods presented in this study.

A great deal of information is provided by a large landmark set between even a single pair of volumetric images. As more patient cases become available, and as the validation feature points are propagated onto the remaining phases of the 4D CT datasets, a more complete and statistically sound characterization of DIR spatial accuracy performance can be achieved. The reference data will be invaluable for optimization of algorithms under development, as the error analysis procedure can be entirely automated to generate formatted error reports as part of the DIR output. This could largely streamline comparative evaluation studies, and allow for more detailed ranking of multi-algorithm spatial accuracy performance that is based on more than simple summary error statistics. With these procedures in place, the problem of formal acceptance testing of a DIR algorithm can be posed as deciding which performance characteristics are most

relevant for the given application, and whether or not the confidence intervals on the measured characteristic errors are acceptable. However, during this process, only those error measurements obtained using patient images equivalent (e.g. 4D CT) to those that will be encountered in clinical practice should be considered.

In order to properly interpret published reports of DIR spatial accuracy for which different reference data sets were utilized, it is important that landmark-based evaluation studies of DIR provide error measurements with clear indication of the observer variance and motion characteristics of the data points, as well as image resolution and voxel dimensions.

#### *2.5.4 Minimum Statistical Requirements on Sample Size*

Using numerical simulation, we demonstrated the statistical uncertainty of the DIR spatial error estimate is inversely proportional to the square root of the number of landmark point pairs and directly proportional to the standard deviation of the spatial error specific to the DIR ( $SD_{DIR}$ ) (figure 2.9). From these statistical considerations and from demonstration of the variation in spatial accuracy with displacement size and anatomic location, we propose that large (>1000) validation landmark sets are indeed necessary for rigorous evaluation of DIR spatial accuracy in the lung.

For comparative evaluation and/or validation of DIR, summary statistics such as mean registration error and standard deviation, should comprise only a component of the overall characterization of DIR spatial accuracy performance, regardless of the sample size from which they are derived. More detailed analyses should be performed investigating the characterization of spatial accuracy with regard to clinically relevant variables that could potentially affect DIR output. This necessarily requires large sample validation data sets and multiple test cases to ensure meaningful statistics for the range of potential clinical variables. In prior studies, a maximum of 108 unique anatomical landmarks, divided between right and left lungs, have been manually identified within a single volumetric image pair for DIR performance assessment [31]. Sarrut et al. have recently reported on a landmark set surpassing 500 feature points distributed over four 4D CT phases and three patients [72]. Using a small number of landmark points, or validation samples restricted to highly selective features, risks under-estimating the mean spatial error and  $SD_{DIR}$ . In this study, we have demonstrated that large (>1100) validation landmark data sets are indeed feasible for rigorous evaluation of DIR spatial accuracy in the lung. To our knowledge, the cases presented here represent the most extensive and

comprehensive set of expert-determined landmark correspondences to date. Efforts are currently underway utilizing the APRIL software to construct a library of manually registered 4D CT data sets, with a requirement of >1000 unique landmarks per case, to facilitate comparative evaluation of DIR for thoracic CT. These data will be made publicly available through our website, <http://www.dir-lab.com>.

#### *2.5.5 Application to Quality Assurance*

Currently, it is not clear that the selection of such large validation landmark sets could ultimately prove feasible for application to routine QA assessment of DIR. However, the presented framework for rigorous evaluation suggests that it may not be necessary either. Ideally, evaluation and characterization of the spatial accuracy performance of a given algorithm should be established prior to clinical acceptance. As mentioned above, those decisions could be based on some evaluation of the performance characteristics that are most relevant for the specific application and whether or not the measured characteristic errors are well-defined and acceptable. Assuming the characterization was based on multiple test cases, each of which consists of relatively large (e.g. >1000) validation landmark sets, one could derive estimates of the landmark sample size necessary to obtain confidence intervals of a specified length and statistical significance about the mean registration error. Thus, only a modest sample size may be necessary to obtain the desired summary error statistics for an arbitrary case. For example, for the two DIR algorithms tested in this study, the sample size requirements for 95% CIs of  $\pm 0.5$  mm are 1050 (OFM) and 36 (MLS). The insight obtained in the prior validation process would be directly applicable for assessing the potential for regional registration errors not necessarily reflected by the sparse set of QA landmarks. In practice, it is likely that a combination QA strategy consisting of landmark point pairs and perhaps some combination of independent evaluations of global DIR performance will prove most effective. For example, Zhong et al. have recently reported on a finite-element based metric for assessing global DIR performance [41]. In that study, the authors propose an automated method for detecting components of the calculated displacement fields that violate principles of continuum mechanics. The concept of unbalanced energy is introduced as an indicator for regions in which the DIR transformation is thought to be of poor quality. Though the proposed method does not provide for direct quantitative assessment of DIR spatial accuracy, it suggests a means for automatically delineating regions in which the DIR performance is suspect. Additional landmark pair locations

could then be weighted towards those suspect regions for local quantitative evaluation. The combination of this type of global assessment with landmark-based measurements of spatial accuracy may prove to be an effective and practical strategy for QA of DIR on a routine clinical basis.

## **2.6 Conclusion**

We have presented a framework and corresponding software infrastructure for rigorous quantitative evaluation of deformable image registration spatial accuracy. The feasibility of generating large (>1100) validation landmark sets has been demonstrated on five component phase pairs from clinically acquired treatment planning 4D CT data. The results demonstrate that large landmark point sets provide an effective means for objective evaluation of DIR with a narrow uncertainty range, and suggest a practical strategy for quality assurance of DIR spatial accuracy on a routine clinical basis.] " [99]



## Chapter 3

### Physiologic Evaluation of Regional Pulmonary Ventilation from Four-Dimensional Computed Tomography

#### Preface

The following material included in this Chapter was originally published in the scientific literature prior to publication of this dissertation document. The original research was conducted as part of this dissertation research, and so is included here as it appears in the published manuscript. The citation for all indented ("[...]") material included in this Chapter is given:

[Richard Castillo, Edward Castillo, Josue Martinez, Thomas Guerrero. Ventilation from Four Dimensional Computed Tomography: Density versus Jacobian Methods. \*Physics in Medicine & Biology\*, \*\*55\*\*: 4661-4685, 2010.](#)

#### 3.1 Introduction

" [Pulmonary ventilation is difficult to measure with physiologically meaningful and quantitative values in three dimensions. Single photon emission computed tomography (SPECT) imaging acquired during administration of a radioactive gas or following administration of a radioactive aerosol is the most widely used clinical imaging method to evaluate ventilation [73-75]. With radioactive aerosols the image is determined by the distribution of radioactivity resulting from the aerosol's distribution during tidal breathing associated with its uptake. Technetium ( $^{99m}\text{Tc}$ )-labeled radiopharmaceutical agent aerosols are known to produce artifacts due to airway deposition, rendering them unsuitable for quantitative use [2, 5]. Xenon ( $^{133}\text{Xe}$ ) dynamic SPECT requires time sequence images of the washout phase after equilibration, from which regional clearance times are calculated [75, 76]. Inert xenon (Xe) gas provides radiographic contrast allowing computed tomography (CT) imaging to measure ventilation from serial images of the wash-in (or wash-out) of Xe into the lung parenchyma using compartmental modeling of the radiographic enhancement [9, 10]. The requirement for temporal sampling of the wash-in phase limits the spatial coverage of this technique to the axial field of view of the CT scanner (<12 cm) and increases the subject's radiation dose. Both of these factors limit its use in human subjects. Positron emission tomography (PET) is another ventilation imaging method (reviewed by Schuster [77]), which uses an inert gas containing a positron emitting isotope, such as nitrogen  $^{13}\text{N}$  [78-81] or neon  $^{19}\text{Ne}$  [82]. Both steady-state and wash-out imaging methods have been described for PET

ventilation imaging. However, these PET isotopes require an on-site cyclotron with a gas delivery system due to the 17.2 s half-life of  $^{19}\text{Ne}$  and 9.97 min half-life of  $^{13}\text{N}$ , limiting their use to research studies. Magnetic resonance (MR) imaging techniques for ventilation have also been developed [11]. With oxygen-enhanced MR imaging, images are acquired before and after a simple change in the inspiratory oxygen concentration; the subtraction images (pure oxygen-breathing image minus the room-air breathing image) are claimed to represent ventilation. Hyperpolarized noble gases, such as xenon ( $^{129}\text{Xe}$ ) or helium ( $^3\text{He}$ ), have been utilized to provide paramagnetic contrast for MR ventilation imaging [83]. MR techniques of ventilation imaging require tracer gases and specialized equipment, which limit the availability of these methods. Additionally, the degree of hyperpolarization is time dependent and the resulting images do not yield quantitative physiological values. A broadly available, inexpensive, quantitative ventilation imaging method is currently lacking.

Four dimensional computed tomography (4DCT) images, developed for radiotherapy treatment planning, also contain CT characteristics that reflect the changes in air content of the lungs due to ventilation. We have developed a method for extracting ventilation images from 4DCT [15] which is potentially better suited and more broadly available for image guided radiotherapy (IGRT) than the current standard SPECT ventilation imaging. 4DCT images are now routinely acquired at many institutions as part of the standard treatment planning for thoracic malignancies [84-90]. Obtaining a ventilation image from the treatment planning 4DCT only requires an additional computational step be performed on the 4DCT images [15]. The change in air content due to ventilation within the lungs provides a radiographic signal representing ventilation. Simon et al. [13] proposed a relationship between manually registered CT values in Hounsfield Units (HU),  $HU_{inhal}$  and  $HU_{exhal}$ , and the regional volume change. Simon's method assumes the fraction of air in a CT region is given by:

$$F_{air} = -\frac{HU}{1000}. \quad (11)$$

The fractional change in air content within a specified volume is the [13]:

$$\frac{\Delta V}{V_{ex}} = \frac{(F_2 - F_1)}{F_1(1 - F_2)}, \quad (12)$$

where  $\Delta V$  is the local volume change due to inspiration,  $V_{ex}$  is the volume of air within the exhalation state volume of interest,  $F_1$  is the fraction of air in the exhale CT volume of interest and  $F_2$  is the corresponding fraction of air in the inhalation state CT volume of interest. If we consider the initial exhale volume as a discretized CT voxel, and further suppose that a three dimensional vector transformation function exists mapping the set of inhale lung CT voxels into the exhale image domain, one can derive the following expression for specific ventilation in terms of registered CT numbers corresponding to inhalation and exhalation breathing states:

$$\frac{\Delta V}{V_{ex}} = 1000 \frac{(HU_{inhale} - HU_{exhale})}{HU_{exhale} (1000 + HU_{inhale})}. \quad (13)$$

In previous work, we used the output from a deformable image registration (DIR) algorithm to link corresponding inhale/exhale volume elements to produce quantitative images of specific ventilation from breath-hold CT (BH-CT) and radiotherapy treatment planning 4D CT [14, 15].

Concerns due to physiological violation of the ventilation model given by equation(13) [15], as well as uncertainty resulting from image noise, acquisition artifacts, and image reconstruction artifacts have led others to investigate alternative methods for quantifying lung function that are independent of the image CT values. Regional volume change may also be calculated from the Jacobian of the DIR deformation result,  $\Phi(\vec{x})$ , alone[7]. Consider  $B_{\vec{x}}$  as a small box around  $\vec{x}$  in the exhale image such that  $\Phi(B_{\vec{x}})$  is the corresponding deformed box in the inhale image. The Jacobian of the deformation  $\Phi(\vec{x})$  at  $\vec{x}$  is:

$$J_{\Phi}(x) = \frac{vol(\Phi(B_{\vec{x}}))}{vol(B_{\vec{x}})}, \quad (14)$$

where  $vol(B_{\vec{x}})$  is the volume of the small box centered at  $\vec{x}$  and  $vol(\Phi(\vec{x}))$  is the volume of the corresponding deformed box [91]. Note the fractional change in volume equals the Jacobian, equation (14). Alternatively, the Jacobian may be calculated from an

analytic formula given by the calculus formulation of the Jacobian operator [91]. The Jacobian calculation is independent of the image CT values, depending only on the DIR transformation function  $\Phi$ .

In this study, we compare two general methodologies for quantifying the physiological specific ventilation from temporal sequences of CT image data acquired without added contrast. The first method ( $sVent_{HU}$ ) is based solely on the physical correspondence of registered HU between image pairs, from which equation (13) may be applied on a voxel-by-voxel basis to yield an image representing quantitative specific ventilation at each voxel position. The resulting ventilation image is inherently co-registered to a common CT frame for spatial correlation with the underlying anatomy. The second methodology utilizes the Jacobian of the DIR spatial transformation to quantify regional expansion and contraction of lung volume elements. In the present work, the quantitative Jacobian calculation is used as an initial measurement from which equivalent co-registered images representing specific ventilation are obtained. Two implementations of the Jacobian-based methodology are investigated, the analytic formulation derived from the definition of the Jacobian operator ( $sVent_{AJ}$ ), as well as the geometric analogue ( $sVent_{GJ}$ ) given by equations (23) and (24). The  $sVent_{GJ}$  is determined by calculating the volume of the tetrahedra that results from displacing the eight voxel vertices from their initial cubic configuration according to the calculated DIR spatial transformation. A quantitative evaluation of these three ventilation calculation methods is performed, first utilizing manually segmented lung parenchymal volumes as a validation metric. This type of analysis has been utilized previously to assess global self-consistency of the calculated volume changes [14, 15, 92]. Additionally, each of the 4DCT-based ventilation image sets will be compared in terms of the spatial overlap of the lowest and highest functioning lung regions, in order to assess the relative distribution of ventilation determined from each method. Finally, similar analyses are performed comparing each of the 4DCT-derived ventilation methods to  $^{99m}Tc$  DTPA aerosol SPECT ventilation images, acquired for each patient on the same day as the 4DCT image sets utilized in this study. The spatial overlap of corresponding percentile ventilation distributions will be assessed for each 4DCT ventilation image relative to the reference SPECT.

### 3.2 Materials and Methods

#### 3.2.1 Thoracic CT Patient Images

All patient images were acquired as part of a 4DCT ventilation validation imaging study, with approval by the M. D. Anderson Institutional Review Board (protocol 2006-0698). Patients with esophagus or lung cancer who were scheduled to receive thoracic radiotherapy were enrolled in this study and all study imaging was performed prior to the initiation of treatment. Image data consists of 7 4DCT data sets, obtained at 2.5 mm slice spacing using a General Electric Discovery ST PET/CT scanner (GE Medical Systems, Waukesha, WI) in the Department of Radiation Oncology at the University of Texas M. D. Anderson Cancer Center. For each of the 7 patients, 4DCT images of the entire thorax and upper abdomen were obtained. Voxel dimensions for each case were  $(0.97 \times 0.97 \times 2.5)$  mm<sup>3</sup>. The 4DCT acquisition technique using the respiratory signal from the Real-Time Position Management Respiratory Gating System (Varian Medical Systems, Palo Alto, CA) has been previously described [93]. The patient and 4DCT image characteristics of the 7 cases utilized in this study are given in table 3.1.

| Case # | Malignancy | Focal Defect | SPECT Image Dimension | SPECT Voxel Dimension (mm) | 4D CT Image Dimension | 4D CT Voxel Dimension (mm) |
|--------|------------|--------------|-----------------------|----------------------------|-----------------------|----------------------------|
| 1      | SCLC       | +            | 128×128×196           | 3.50×3.50×2.00             | 512×512×128           | 0.97×0.97×2.50             |
| 2      | Eso ca.    | -            | 128×128×196           | 2.81×2.81×2.00             | 512×512×136           | 0.97×0.97×2.50             |
| 3      | NSCLC      | -            | 128×128×153           | 2.87×2.87×2.50             | 512×512×128           | 0.97×0.97×2.50             |
| 4      | Eso ca.    | -            | 128×128×196           | 2.34×2.34×2.00             | 512×512×128           | 0.97×0.97×2.50             |
| 5      | NSCLC      | -            | 128×128×153           | 2.90×2.90×2.50             | 512×512×120           | 0.97×0.97×2.50             |
| 6      | NSCLC      | +            | 128×128×196           | 2.81×2.81×2.00             | 512×512×120           | 0.97×0.97×2.50             |
| 7      | NSCLC      | +            | 128×128×153           | 2.74×2.74×2.50             | 512×512×136           | 0.97×0.97×2.50             |

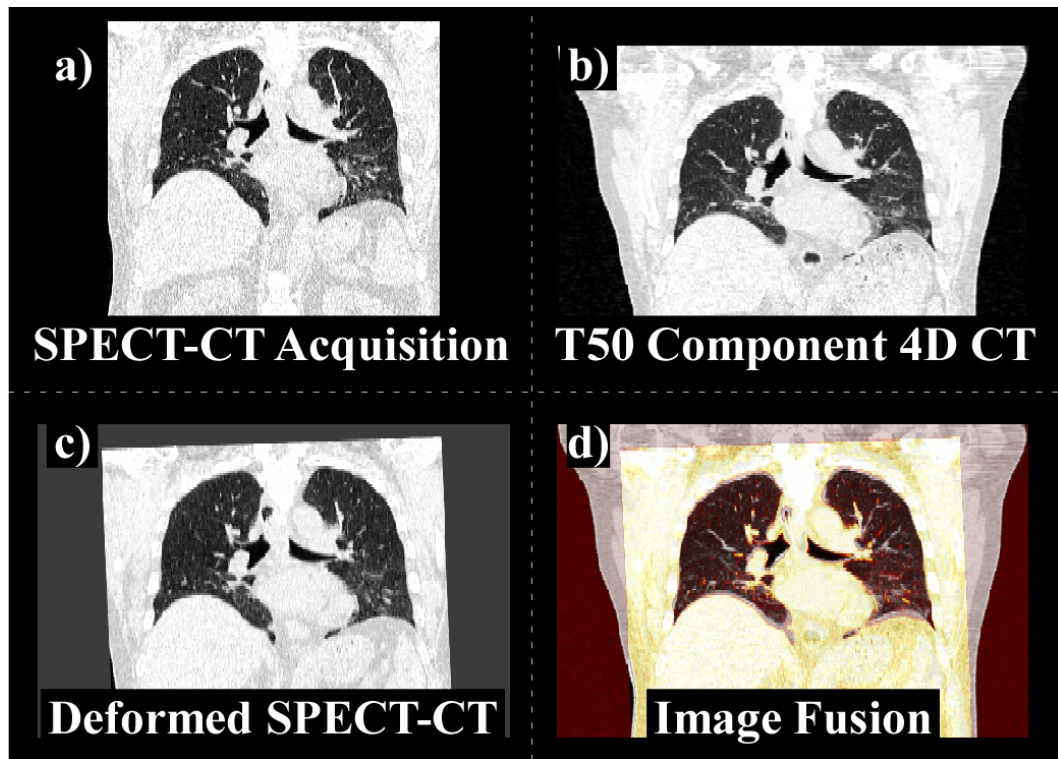
abbreviations: SCLC = small cell lung cancer; NSCLC = non-small cell lung cancer; Eso ca. = esophagus cancer

**Table 3.1. Ventilation cases.** The patient data included in this study is summarized in the table above. The presence of focal ventilation defects was determined clinically from the <sup>99m</sup>Tc-DTPA aerosol SPECT ventilation exams, and recorded in each patient's medical record. The SPECT descriptors refer to the raw nuclear medicine exam data, prior to affine registration with 4D CT.

#### 3.2.2 Pulmonary Ventilation SPECT Imaging

Each patient also received ventilation SPECT imaging on a Siemens Simbia T6 SPECT-CT scanner (Hoffman Estates, IL) on the same day as the 4DCT acquisition. A sub-micronic <sup>99m</sup>Tc-diethylenetriamine-pentaacetic acid (<sup>99m</sup>Tc-DTPA) aerosol was generated by jet nebulization (Medi-Physics, Arlington Hts, IL) and administered to the resting tidal

breathing patients.  $^{99m}\text{Tc}$ -DTPA aerosol was administered while the patients were seated with up-right posture prior to acquisition of the emission image. A co-registered CT image was also acquired to perform attenuation correction of the  $^{99m}\text{Tc}$  140 keV emission photons. The volumetric CT was also used for anatomical reference and for affine registration with the maximum exhalation component phase image from the corresponding 4DCT data set, using the CT-to-CT fusion software in a commercial treatment planning system (Pinnacle, version 8.1x, Philips Medical Systems, Andover, MA). The affine coefficient matrix for each case was subsequently used to align the SPECT and 4DCT-derived ventilation images for comparative evaluation. The use of affine registration for functional and anatomical alignment has been previously reported in radiotherapy treatment planning studies requiring multi-modality image fusion or spatial registration of anatomical and quantitative functional image information [16, 94-97]. Figure 3.1 illustrates the alignment process for an example case utilized in this study.



**Figure 3.1. CT-to-CT affine registration.** **a)** A coronal section is shown through the CT component of the SPECT-CT exam for an example case (#2). **b)** A similar coronal section is shown through approximately the same anterior-posterior position in the maximum exhalation phase image from the corresponding 4D CT. **c)** The coronal section from (a) is shown following

affine registration with the 4D CT depicted in (b). **d)** A fusion image is shown demonstrating the alignment of the SPECT-CT and 4D CT data sets. The aligned image from (c) is shown in yellow. The affine coefficient matrix was determined using the CT-to-CT fusion software in a commercially available treatment planning system (Pinnacle, version 8.1x, Philips Medical Systems, Andover, MA).

### 3.2.3 4DLTM Deformable Image Registration

We developed a four-dimensional spatio-temporal image registration (4D DIR) algorithm, which links all expiratory phases using four-dimensional local trajectory modeling (4DLTM) [98]. In this study, the cubic 4DLTM algorithm was applied to the six expiratory phase images from each 4DCT set, beginning with the maximum inhalation phase (T00) and ending on the maximum exhalation phase (T50). We previously reported the spatial accuracy of the 4DLTM algorithm using the method described by Castillo et al. [99]. All ventilation images were subsequently derived using the same 4DLTM output for each ventilation calculation method.

### 3.2.4 Lung Parenchyma Segmentation

An intensity-based segmentation algorithm was applied to delineate lung voxels from the CT images, with CT values in the range [-999, -250] HU selected as representing pulmonary parenchyma. The trachea and main-stem bronchi were separately delineated by three-dimensional morphological growing [63] from initial seed locations, and subsequently removed from the pulmonary parenchyma masks. A single seed point grown from the center of the heart was similarly used to extract pulmonary vasculature structures from the lung mask. The binary mask images were subsequently used to define the spatial domain for lung function quantification, as well as self-consistency analyses as described in sections 3.2.5-8.] "

### 3.2.5 Density-based Specific Ventilation from 4D CT

Voxel-wise quantification of regional ventilation from CT image intensities requires a spatial correspondence of HUs between pairs of respiratory phase images. Consider the voxel located at position  $\vec{x}_{inhale}$  in the inhalation state image. Define the spatial transformation function  $\Phi$  such that the corresponding position of  $\vec{x}_{inhale}$  in the exhalation state image is given by:

$$\vec{x}_{exhale} \equiv \Phi(\vec{x}_{inhale}) = \vec{x}_{inhale} + T(\vec{x}_{inhale}), \quad (15)$$

where  $T$  is the vector displacement field associated with the spatial transformation  $\Phi$ . Note that  $\Phi$  must be defined as such, so that expanded tissue elements can be mapped back into their original location in the compressed (exhale) state. Given  $\Phi$ , one can derive the following expression for specific ventilation in terms of registered CT numbers corresponding to inhalation and exhalation breathing states, where the resultant function image is inherently co-registered with the anatomical exhalation state CT (derivation provided in Appendix):

$$\frac{\Delta V}{V_{exhale}^{air}} = 1000 \frac{(\bar{H}_{inhale} - HU_{exhale})}{HU_{exhale} (1000 + \bar{H}_{inhale})}, \quad (16)$$

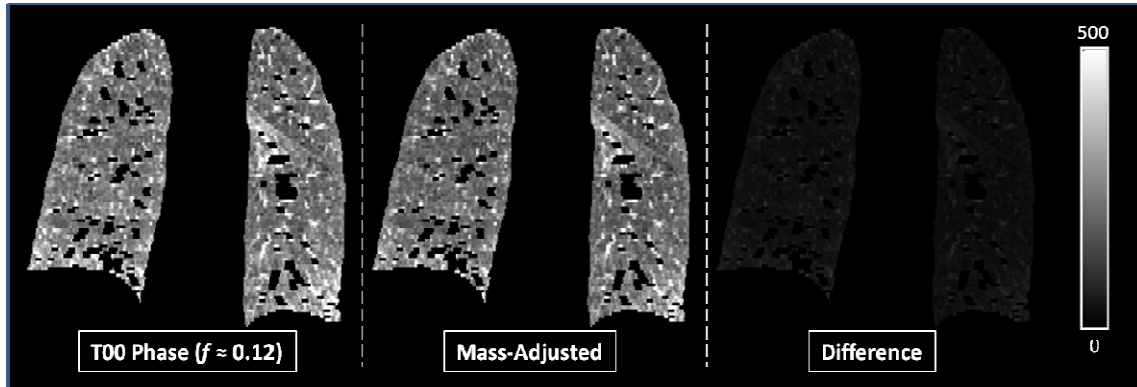
where  $\bar{H}_{inhale}$  represents the mean HU of the tissue volume that is mapped into the exhale voxel with corresponding CT number given by  $HU_{exhale}$ . The quantity  $(\Delta V/V_{exhale}^{air})$  is the fractional volume change due to inspiration, i.e., the local specific ventilation. Equation (16) differs slightly from the previous formulation (equation (13)) originally reported by Guerrero et al. in 2005 [14], in that all instances of the quantity  $HU_{inhale}$  have been replaced with  $\bar{H}_{inhale}$ . The subtle difference is a necessary modification to account for the full spatial extent of the expanded tissue element in the inhalation state. In general, inflation of a single tissue element in the exhalation state image will result in its inflated volume spanning multiple volume elements in the inhalation state image. It can be shown that the specific ventilation at the initial exhale voxel position is related, rather to the average HU comprising the expanded volume in the inhale image. Including only a single inhale HU fails to incorporate the physiologic information encoded in the remaining CT numbers corresponding to the same initial tissue element, and thus is not strictly correct.

Previously, it has been shown that respiratory-induced changes in tissue perfusion result in cyclic variation in the apparent lung mass determined from 4DCT [15]. The contribution of tissue perfusion to the lung CT numbers is a subtle violation of the basic principles from which equation (16) is derived. Therefore, a correction is first applied to each lung voxel in the inspiration state image to account for the observed difference in CT-derived mass between the two images:

$$HU_{inhale}^{corrected} = HU_{inhale}^{observed} - 1000 f \left( 1 + \frac{HU_{inhale}^{observed}}{1000} \right), \quad (17)$$



where  $f$  is the observed fractional discrepancy in total lung mass. Equation (17) is itself an approximation, which assumes uniform effect of tissue perfusion on the global lung density (figure 3.2). However, previously it has been applied and shown to yield quantitative ventilation images that are significantly better correlated with the measured tidal volume in both human [14-16] and animal studies [92, 100]. Equations (16) and (17) are then applied at each exhale voxel position defined by the lung parenchyma mask. The result is a parametric map, co-registered with the exhalation CT, in which voxel intensities represent fractional change in air content within the originating compressed sub-volume that occurred due to inspiration. In this study, local  $3 \times 3 \times 3$  voxel averages were used for each HU input into equation (16) to account for noise in the CT image intensities, as well as small spatial errors in the DIR. The local HU averages were masked by the lung parenchyma segmentation images in order to avoid implicit inclusion of non-lung voxels into the ventilation calculations. The final three-dimensional ventilation images were created following smoothing with  $9 \times 9 \times 3$  voxel box average filter.



**Figure 3.2. Mass correction for inhale tissue elements.** A correction is applied to each voxel in the inspiration state image to account for the effect of tissue perfusion on the observed image intensities. The correction is a function of the observed mass discrepancy, and assumes the perfusion effect is a uniform increase in soft tissue density throughout the lung (i.e.,  $HU \approx HU + \Delta HU$ ).

### 3.2.6 Jacobian-based Specific Ventilation: Analytic Implementation

" [Regional compression or expansion of image sub-volumes can also be determined directly from the DIR spatial transformation and subsequently used to derive equivalent measures of the physiological specific ventilation. In this study, both analytic and geometric implementations of the Jacobian operator are utilized to extract volume

changes on a voxel level directly from the vector valued function  $\Phi$  given by the DIR. For the analytic implementation, the determinant of the Jacobian matrix is calculated at each inhale voxel position  $\bar{x}_{inhale}$ , according to:

$$J(\bar{x}_{inhale}) = \det \begin{bmatrix} 1 + \frac{\partial T_1(\bar{x}_{inhale})}{\partial x_1} & \frac{\partial T_1(\bar{x}_{inhale})}{\partial x_2} & \frac{\partial T_1(\bar{x}_{inhale})}{\partial x_3} \\ \frac{\partial T_2(\bar{x}_{inhale})}{\partial x_1} & 1 + \frac{\partial T_2(\bar{x}_{inhale})}{\partial x_2} & \frac{\partial T_2(\bar{x}_{inhale})}{\partial x_3} \\ \frac{\partial T_3(\bar{x}_{inhale})}{\partial x_1} & \frac{\partial T_3(\bar{x}_{inhale})}{\partial x_2} & 1 + \frac{\partial T_3(\bar{x}_{inhale})}{\partial x_3} \end{bmatrix}. \quad (18)$$

Without loss of generality, we assume an initial voxel volume of unity, such that both the total volume change as well as the specific volume change,  $sVol^*$ , at each inhale voxel position are given simply by:

$$sVol_{AJ}^*(\bar{x}_{inhale}) = 1 - abs[J(\bar{x}_{inhale})]. \quad (19)$$

Note that this quantity differs from the physiologic specific ventilation given above, since equations (18) and (19) yield specific volume change relative to the initial voxel volume itself, rather than the initial gas volume within the voxel. Strict application of equation (19) on a voxel-by-voxel basis thus yields a quantitative image of the specific volume change, inherently co-registered with the inhalation state component image. Though the resulting image contains the set of information regarding local voxel expansion and contraction derived from the CT image pair, it is itself not a proper depiction of the spatial distribution of ventilation. This is because the parametric map itself does not reflect the spatial correspondence of expanded sub-volumes with the initial compressed state. Thus, in order to generate the final quantitative and spatially configured functional image, all local voxel measurements of volume change must first be accumulated at each originating voxel position in the exhale image.

To perform the accumulation step, consider again the voxel  $\bar{x}_{inhale}$  in the inhale lung image. Define  $\bar{X}_{exhale}$  as the corresponding voxel in the exhale image that is linked by the spatial transformation  $\Phi$ , such that  $\Phi(\bar{x}_{inhale}) = \bar{X}_{exhale}$ . Since, in general, local

expansion will result in multiple such inhale voxels similarly being mapped into  $\vec{X}_{exhale}$ , we define  $\mathbf{M}$  as the set of all such inhale voxel positions:

$$\mathbf{M}(\vec{X}_{exhale}) = \{\vec{x}_i \in Inhale \mid \Phi(\vec{x}_{inhale}) = \vec{X}_{exhale}\}. \quad (20)$$

Note that for a single DIR solution  $\Phi$ , the set  $\mathbf{M}(\vec{X}_{exhale})$  is consistent for each specific ventilation formulation. The total fractional change in air content at position  $\vec{X}_{exhale}$  is then given by the sum of the individual fractional changes calculated between  $\vec{X}_{exhale}$  and the set of corresponding inhale voxels given in  $\mathbf{M}$ . Thus, the spatially configured specific volume change at each exhale voxel position  $\vec{X}_{exhale}$  is given by:

$$sVol_{AJ}(\vec{X}_{exhale}) = \sum_{\mathbf{M}(\vec{X}_{exhale})} sVol_{AJ}^*(\vec{x}_i). \quad (21)$$

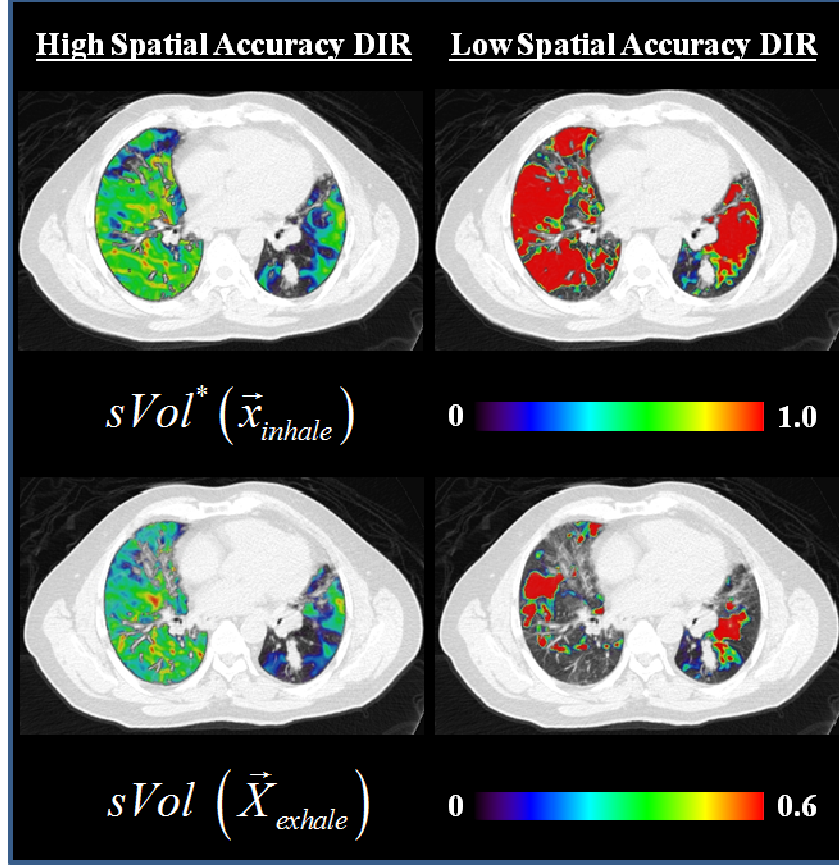
Figure 3.3 illustrates the effect of proper spatial configuration on the calculated parametric map, using “high” and “low” spatial accuracy DIR solutions as examples demonstrating the impact of labeling each exhale position according to its *total* calculated change in volume due to inspiration. Visualization of the spatially configured functional images shows spatial localization of both over- and under-represented pulmonary regions in the “low” spatial accuracy DIR.

Once co-registered with the exhalation state image, voxel  $\vec{X}_{exhale}$  in the Jacobian-derived image of specific volume change is converted into specific ventilation by dividing the total calculated volume change at position  $\vec{X}_{exhale}$  by the initial exhale gas volume at that location, given simply by equation (11):

$$sVent_{AJ}(\vec{X}_{exhale}) = sVol_{AJ}(\vec{X}_{exhale}) * \left[ -\frac{HU(\vec{X}_{exhale})}{1000} \right]^{-1}. \quad (22)$$

Note that although regional expansion and contraction is determined directly from the calculated DIR, the exhale HU are necessary in equation (22) to quantify the physiologic

specific ventilation, since they provide direct measurement of the initial gas volume within each discrete volume element. All image derivatives required by equation (18) were calculated using forward finite difference approximations [101] applied to the image grid.



**Figure 3.3. Spatial configuration of specific volume change.** Each measurement of Jacobian-derived volume change must be registered with the corresponding originating sub-volume. In this way, each exhale position is labeled according to its total change in volume due to inspiration.

### 3.2.7 Jacobian-based Specific Ventilation: Geometric Implementation

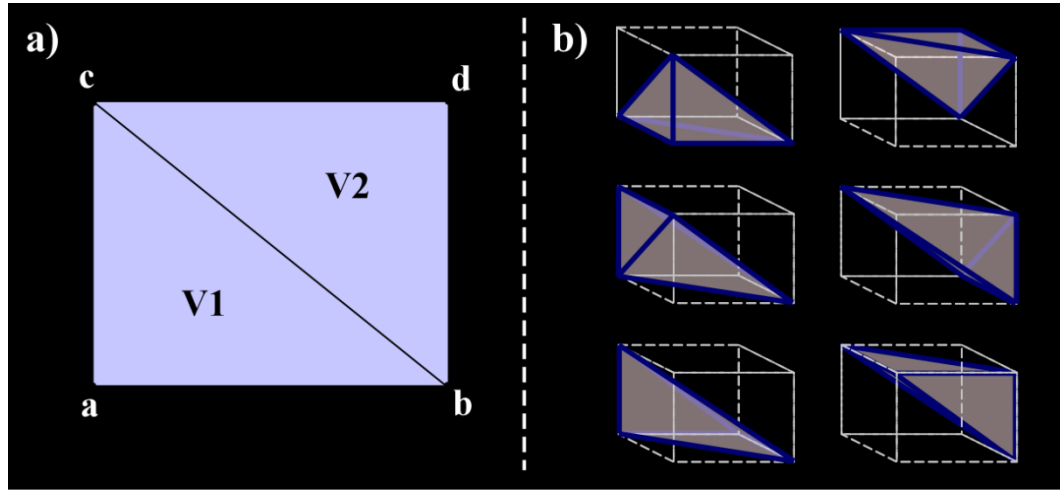
For the geometric implementation of the Jacobian calculation, specific volume change is conceptually obtained by applying the calculated DIR spatial transformation  $\Phi$  to each of the eight voxel vertex positions and subsequently calculating the volume of the deformed volume element. In practice, this is accomplished by first partitioning each voxel into six independent tetrahedra (figure 3.4). Let  $\vec{a}, \vec{b}, \vec{c}$  and  $\vec{d}$  correspond to the

three-dimensional vector coordinates of the four vertices of a single tetrahedron face in the inhalation state CT image at voxel position  $\bar{x}_{inhale}$ . The magnitude volume of the  $i^{th}$  tetrahedron, deformed according to  $\Phi$ , is given by:

$$V_i(\bar{x}_{inhale}) = \frac{1}{3!} \det \begin{bmatrix} \Phi(a_i)_1 & \Phi(a_i)_2 & \Phi(a_i)_3 & 1 \\ \Phi(b_i)_1 & \Phi(b_i)_2 & \Phi(b_i)_3 & 1 \\ \Phi(c_i)_1 & \Phi(c_i)_2 & \Phi(c_i)_3 & 1 \\ \Phi(d_i)_1 & \Phi(d_i)_2 & \Phi(d_i)_3 & 1 \end{bmatrix}, \quad (23)$$

where the subscripts 1, 2, and 3 correspond to the respective  $x$ ,  $y$ , and  $z$  spatial components of the vector valued function  $\Phi$ . The final volume of the deformed initial voxel  $\bar{x}_{inhale}$  is then given by the sum of the volumes of the six deformed tetrahedra:

$$V(\bar{x}_{inhale}) = \sum_{i=1}^6 V_i(\bar{x}_{inhale}). \quad (24)$$



**Figure 3.4. Specific volume change via the geometric Jacobian implementation.** The geometric Jacobian (GJ) method for computing the volume of a deformed volume element is based on subdividing an initial voxel configuration with unit volume into six independent tetrahedra. **a)** In two dimensions, this procedure simplifies to the splitting of a unit area pixel in two triangles with vertices  $(\vec{a}, \vec{b}, \vec{c})$  and  $(\vec{d}, \vec{c}, \vec{b})$ . **b)** The three-dimensional voxel partition is shown schematically for comparison.

As with the analytic Jacobian implementation, specific volume change relative to the original voxel volume for the geometric implementation is given by:

$$sVol_{GJ}^* (\vec{x}_{inhale}) = 1 - V(\vec{x}_{inhale}). \quad (25)$$

Conversion to quantitative and spatially configured specific ventilation is subsequently achieved by accumulating individual volume changes back into the exhale image frame and dividing by the initial gas volume. Assuming the same DIR transformation  $\Phi$ , the spatially configured specific volume change at each exhale voxel position is given by:

$$sVol_{GJ}(\vec{X}_{exhale}) = \sum_{M(\vec{X}_{exhale})} sVol_{GJ}^*(\vec{x}_i). \quad (26)$$

Specific ventilation is similarly obtained:

$$sVent_{GJ}(\vec{X}_{exhale}) = sVol_{GJ}(\vec{X}_{exhale}) \left[ -\frac{HU(\vec{X}_{exhale})}{1000} \right]^{-1}. \quad (27)$$

Both density change- and Jacobian-based specific ventilation images derived in this study were smoothed with a 9×9×3 voxel box average filter to equivalently yield the final quantitative functional map, co-registered with the maximum exhalation component phase image.

### 3.2.8 Comparative Evaluation of Ventilation Methodologies

Evaluation of the CT-derived ventilation images was assessed in terms of self-consistency, in addition to quantitative comparison with the clinical standard  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation.

For self-consistency analysis, separate right and left lung tidal volumes were derived using the CT ventilation images, and compared to direct measurements obtained from the corresponding segmented lung parenchyma mask images. The volume of each segmented lung was subsequently quantified and the difference between corresponding

maximum inhale and exhale image pairs, representing the measured tidal volume, used for comparison with that determined directly from the ventilation images.

In addition to comparative evaluation of the magnitude specific ventilation, the relative distribution of ventilation was also quantitatively assessed and compared among methods, as well as with the clinical standard  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation. To assess the relative distribution of lung function, each of the SPECT and 4DCT-derived ventilation images was mapped into corresponding percentile images in which each voxel value was replaced by the density of the voxel intensity given by the corresponding cumulative distribution function. For the SPECT ventilation data sets, the percentile images were derived only from the set of SPECT image intensities included in the lung parenchyma segmentations used to mask each of the 4DCT-derived ventilation calculations.

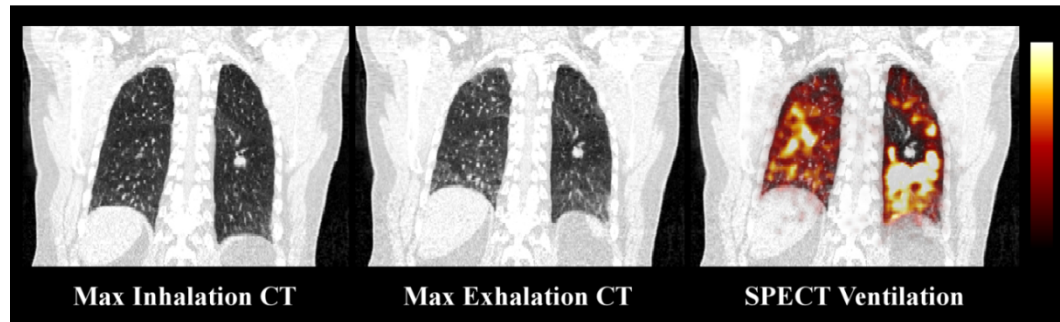
Inherently, the  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation images are not strictly quantitative. Ultimately, their utility arises in the ability to localize regional ventilation deficits, such as arising from airway obstruction (see figure 3.5) or chronic obstructive pulmonary disease (COPD), which is potentially useful in radiotherapy treatment planning for IGRT. Though the CT-derived measures of specific ventilation are inherently quantitative and bearing physiological significance at a voxel level, the use of quantitative  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation is not particularly relevant for purposes of validation or comparative evaluation, since the quantitative component itself is not clinically constructive. A more meaningful assessment is in the ability to identify and spatially delineate regions of poor pulmonary function, in order to determine how those regions correlate spatially between methods. Since there is the potential that intensity artifacts due to airway deposition of the  $^{99m}\text{Tc}$  DTPA aerosol will mask local foci of functional deficit, evaluation should only be over those regions where the SPECT ventilation images clearly indicate regional functional loss, such as indicated by the nuclear medicine physician in the formal clinical record. Note that in general this is not necessarily the lowest percentile ventilation in each image, but rather the lowest percentile ventilation in cases which have a marked and recorded functional deficit. Additionally, the percentile cut-off for delineating functional loss is directly related to the spatial extent within the effected lung, which will certainly vary among cases. Thus, to provide some quantitative indication as to the ability of the 4DCT-based ventilation images to delineate macroscopic foci of ventilation deficit relative to the clinical reference, each of the  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation images was partitioned

into non-overlapping regions representing fixed intervals of percentile distribution of ventilation. Percentile regions were selected in 20% intervals, resulting in binary masks for each case representing the spatial distribution of (1-20), (41-60), (61-80), and (81-100)% percentile regions within the segmented lung volume. Corresponding mask regions were similarly delineated using both HU- and Jacobian-derived ventilation images. The quantitative Dice similarity coefficient (DSC) between SPECT and each of  $sVent_{HU}$ ,  $sVent_{AJ}$ , and  $sVent_{GJ}$  percentile segments was determined for all cases. The DSC is a measure of the degree of overlap between two areas or volumes, and is quantified as the ratio of twice the volume of intersection to the sum of the two volumes [102, 103]. To assess the statistical difference among methods, we fit a three factor ANOVA model to the DSC outcomes, where the method is a three level factor consisting of  $DSC_{HU}$ ,  $DSC_{AJ}$ , and  $DSC_{GJ}$ . The remaining two factors are the percentile mask, which is composed of the five levels (1-20), (41-60), (61-80), and (81-100)%, and the 7 patients included in the study. The statistical differences were evaluated using the Tukey multiple comparison procedure. Additionally, the DSC was similarly determined for all combinations of 4DCT-based ventilation methods.

### 3.3 Results

#### 3.3.1 4D CT and $^{99m}Tc$ Aerosol SPECT Image Properties

Table 3.1 shows a summary of the clinical characteristics of the patient data set included in this study, as well as image properties for the 4D CT and SPECT ventilation exams. The presence of macroscopic ventilation defects reported by the nuclear medicine physician is also shown for the set of seven cases.



**Figure 3.5. 4D CT and SPECT ventilation.** a) A maximum inspiration phase 4D CT image is shown in coronal section. b) The corresponding section is shown for the maximum expiration phase 4D CT image. c) The  $^{99m}Tc$ -DTPA aerosol SPECT-CT was registered to the maximum



expiration phase CT image using affine registration. A corresponding coronal section is shown superimposed.

| Case # | Inhale<br>Vol (mL) | Exhale<br>Vol (mL) | Tidal<br>Vol (mL) | Avg (SD)<br>Displ (mm) | Inhale<br>Avg (HU) | Exhale<br>Avg (HU) |
|--------|--------------------|--------------------|-------------------|------------------------|--------------------|--------------------|
| 1      | 2137               | 1612               | 525               | 9.5 (5.0)              | -738               | -669               |
| 2      | 3076               | 2394               | 682               | 10.1 (6.4)             | -788               | -749               |
| 3      | 4970               | 4044               | 926               | 13.6 (8.3)             | -851               | -826               |
| 4      | 1488               | 1206               | 282               | 6.7 (3.5)              | -722               | -674               |
| 5      | 2600               | 2133               | 467               | 7.9 (5.7)              | -826               | -802               |
| 6      | 3462               | 2826               | 636               | 12.7 (11.0)            | -771               | -745               |
| 7      | 3416               | 2954               | 462               | 4.7 (3.8)              | -855               | -840               |

abbreviations: SD = standard deviation; HU = Hounsfield Unit

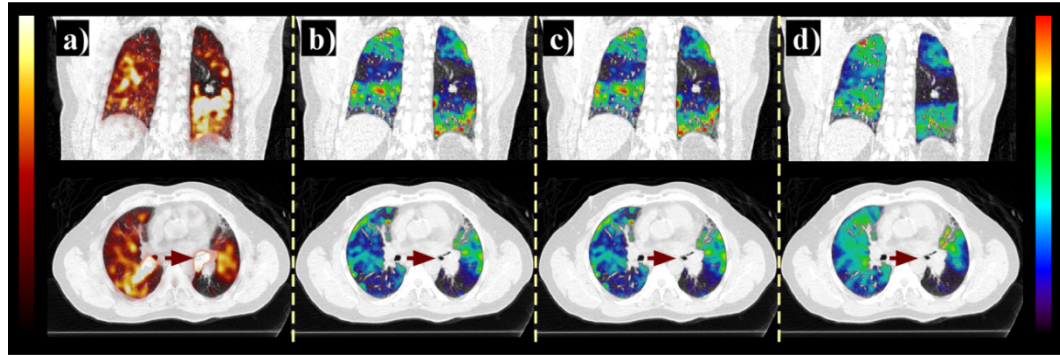
**Table 3.2. 4D CT characteristics.** Functional measurements obtained from the segmented lung mask regions. Average displacements were determined from the calculated 4D DIR displacements within the lung masks.

Figures 3.5a & b show coronal images from the extreme inspiration and expiration phases of a typical 4DCT image set (case #1 in this study). The lung volumes, tidal volume, and average CT value in HU were determined from the segmented CT images, with measured values reported in table 3.2. The 4DLTM DIR was applied to the expiratory phases linking the images from the extreme inhale to the extreme exhale phases. The average lung voxel displacements were obtained directly from the calculated 4DLTM transformation functions and are also reported in table 3.2. Tidal volumes ranged from 282 to 926 mL, while the average calculated lung voxel displacement ranged from 4.74 to 13.55 mm. The corresponding  $^{99m}\text{Tc}$ -DTPA aerosol SPECT image is shown in figure 3.5c superimposed on the extreme expiration phase CT image, following affine registration as described in section 3.2.2. Note the presence of a ventilation defect in the left lung and the high uptake artifacts resulting from  $^{99m}\text{Tc}$ -DTPA aerosol deposition in the airway.

### 3.3.2 CT-derived Specific Ventilation

4DCT-derived specific ventilation images obtained using the three experimental methods ( $s\text{Vent}_{\text{AI}}$ ,  $s\text{Vent}_{\text{GI}}$ , and  $s\text{Vent}_{\text{HU}}$ ) described in sections 3.2.5-7 are shown in figure 3.6. The sample case shown is one in which gross tumor involvement has caused narrowing of the left main bronchus, resulting in regional hypo-ventilation that is visible on the  $^{99m}\text{Tc}$ -

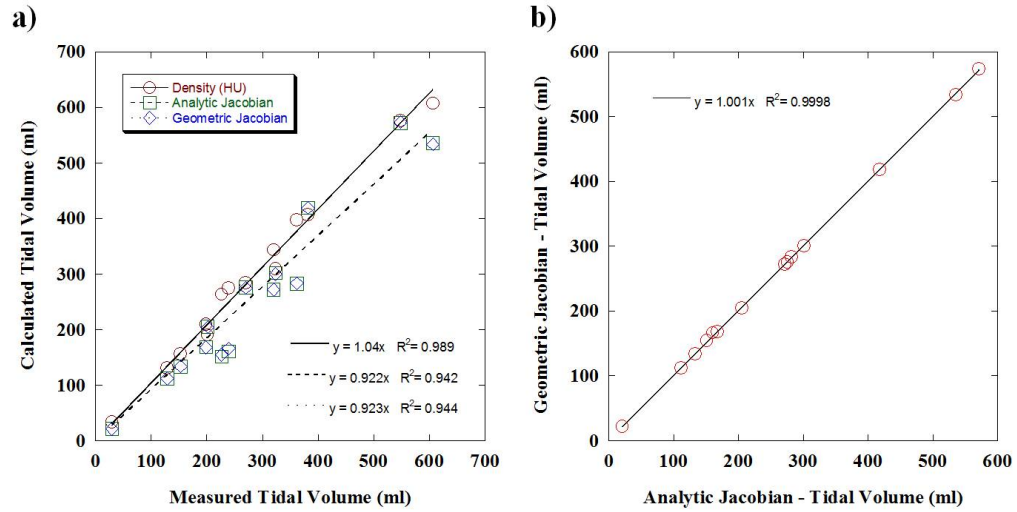
DTPA aerosol SPECT ventilation image (figure 3.6a), and documented in the clinical nuclear medicine report (see table 3.1). The images in figures 3.5 & 3.6 are from this case and represent the same coronal plane. Regional heterogeneity is apparent for each of the 4DCT-derived ventilation images, though the spatial distribution is visibly different between Jacobian- and density change-based methodologies. This is particularly evident in the transverse sections shown in the bottom panel of figure 3.6.



**Figure 3.6 SPECT and 4D CT ventilation.** **a)** The  $^{99m}\text{Tc}$ -DTPA aerosol SPECT-CT was registered to the maximum expiration phase 4D CT image; an arrow is shown indicating bronchial obstruction by the gross tumor volume. **b)** Ventilation image calculated from a 4D CT set using the geometric Jacobian implementation ( $\text{sVent}_{\text{GJ}}$ ), **c)** the analytic Jacobian implementation ( $\text{sVent}_{\text{AJ}}$ ), and **d)** the density-based ventilation ( $\text{sVent}_{\text{HU}}$ ) are shown superimposed on the maximum expiration phase 4D CT image.

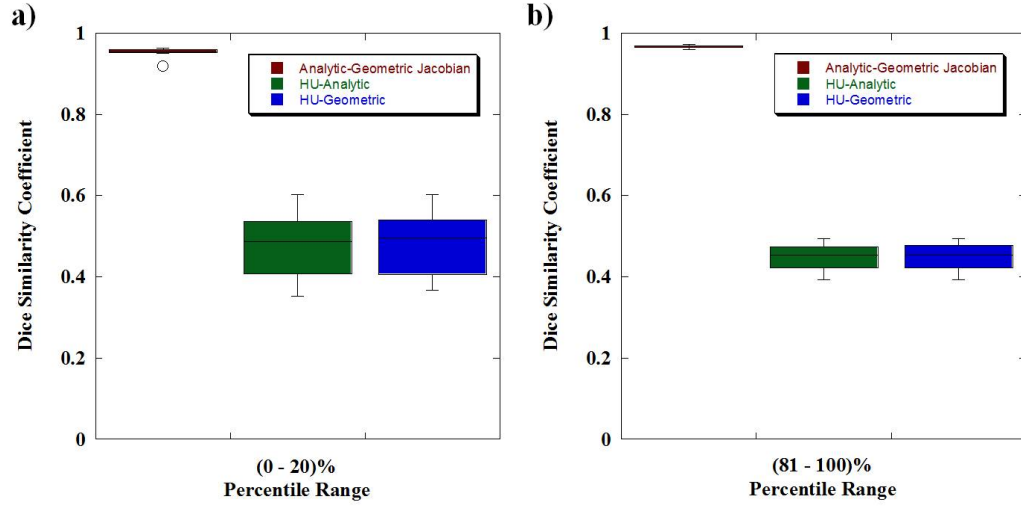
Figure 3.7a shows a comparison of calculated tidal volumes obtained directly from the specific ventilation images with measured values derived from the volumes of the lung parenchyma binary masks. Note that right and left lung masks were treated separately, so as to provide more data points for self-consistency evaluation. Linear regression analysis resulted in slopes of 1.04 ( $R^2 = 0.99$ ), 0.92 ( $R^2 = 0.94$ ), and 0.92 ( $R^2 = 0.94$ ) for respective  $\text{sVent}_{\text{HU}}$ ,  $\text{sVent}_{\text{AJ}}$ , and  $\text{sVent}_{\text{GJ}}$  ventilation images. Thus, on a global scale, both density- and Jacobian-based methodologies yielded good agreement with image segmentation-based measures of resting tidal volume. Figure 3.7b shows direct comparison of calculated tidal volumes derived from analytic and geometric implementations of the Jacobian-based methodology. Linear regression analysis (slope = 1.001,  $R^2 = 0.99$ ) demonstrates equality of the two calculations on a global scale,

providing further indication that they are merely two numerical implementations of a single underlying formulation.



**Figure 3.7. 4D CT-derived tidal volumes.** **a)** Measured versus calculated tidal volumes are shown for each of the 4D CT-based ventilation imaging methods presented. **b)** Direct comparison of calculated tidal volumes for both implementations of the Jacobian-based ventilation imaging methods. Linear regression demonstrates equality of the two calculations, suggesting they are merely two numerical implementations of a single underlying formulation.

Quantitative regional comparisons of the relative spatial distribution of ventilation within the percentile mask regions are shown in the box plots in figure 3.8 for all combinations of 4DCT-derived ventilation methods. The figure shows DSC values over the clinically relevant percentile distribution ranges representing the lowest (figure 3.8a) and highest (figure 3.8b) 20<sup>th</sup> percentile ventilation. Figure 3.8 reflects the regional consistency of the global equality between Jacobian-based methods that is illustrated in figure 3.7b. Furthermore, it is evident from the figure that the density change- and Jacobian-based methods are not equivalent for delineating the highest and lowest functioning lung regions.



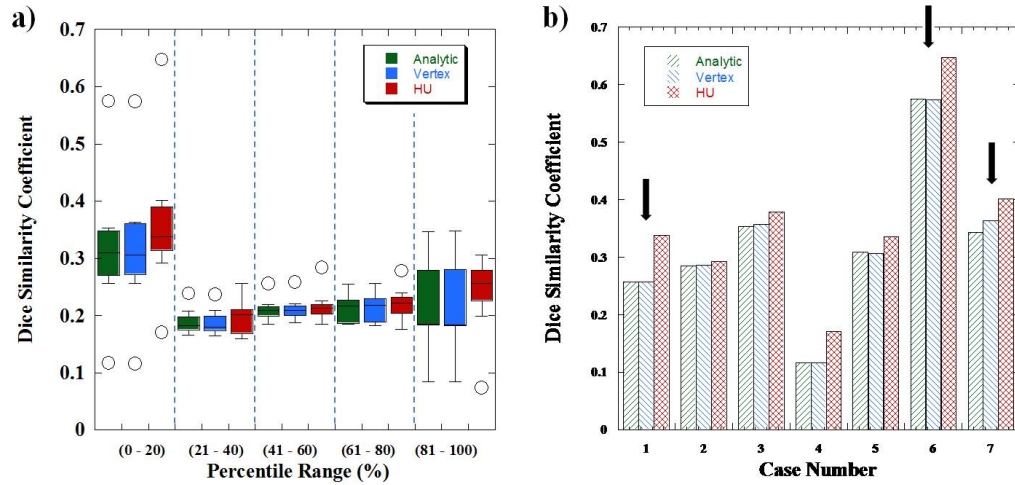
**Figure 3.8. 4D CT percentile ventilation distributions.** The Dice similarity coefficient between all combinations of 4D CT-derived percentile ventilation distributions. The figure quantifies the volumetric overlap between methods among both the **a)** lowest (0-20)% and **b)** highest (81-100)% percentile ventilation spatial distributions.

### 3.3.3 Comparison with $^{99m}\text{Tc}$ DTPA Aerosol SPECT Ventilation

Figure 3.9a shows a boxplot comparison of DSC values for all 4D CT-based ventilation images over all percentile mask regions. We find a significant difference in DSC between the three methods, with  $\text{sVent}_{\text{HU}}$  having the greatest DSC when compared to both Jacobian-based implementations, with ( $p < 10^{-4}$ ). No significant difference was found between  $\text{sVent}_{\text{AJ}}$  and  $\text{sVent}_{\text{GJ}}$ , with ( $p = 0.92$ ). We also find significant differences between percentile mask regions, with the (1-20)% segment having the largest DSC when compared to all others ( $p < 10^{-4}$ ). Finally, we note that there is a string interaction between patient and percentile mask regions ( $p < 10^{-4}$ ), and a noticeable interaction between percentile mask and ventilation method factors ( $p = 0.0158$ ). Analysis of the residuals (not shown) demonstrated that the statistical model was appropriate for the experimental design presented. Figure 3.9a clearly illustrates that the highest spatial correlation between the set of 4D CT-derived ventilation images with the  $^{99m}\text{Tc}$ -DTPA aerosol SPECT ventilation occurred in the lowest function percentile mask region.

Figure 3.9b shows raw DSC values in the (1-20)% percentile range for the set of 7 patient cases. Arrows are shown indicating the 3 cases for which marked ventilation deficits were recorded by the nuclear medicine physician as part of the clinical  $^{99m}\text{Tc}$

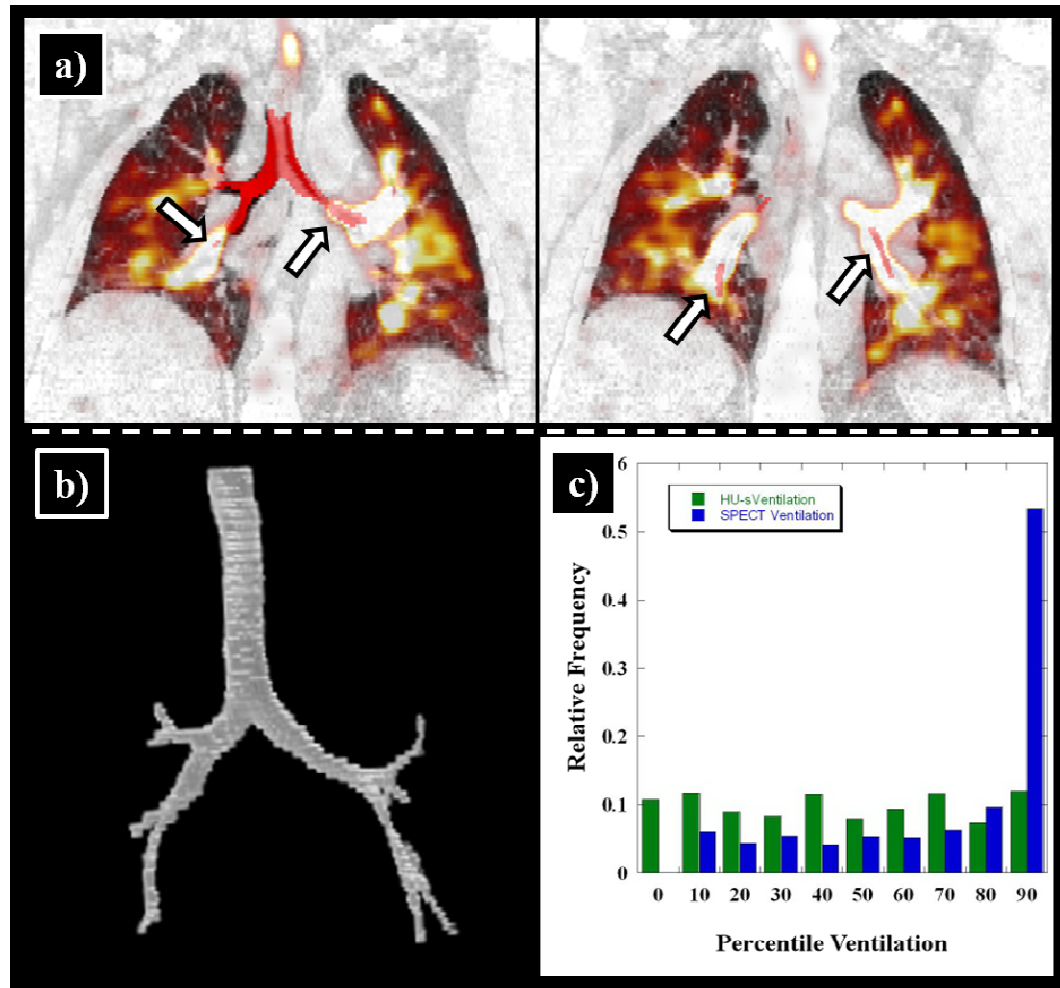
DTPA aerosol SPECT ventilation exam (see table 3.1). The figure suggests that the statistical difference achieved in Figure 3.9a between  $sVent_{HU}$  and Jacobian-based methods is the result of better local correlation between  $sVent_{HU}$  and the clinical  $^{99m}Tc$  DTPA aerosol SPECT ventilation in regions of marked ventilation deficit.



**Figure 3.9. Dice similarity evaluation.** **a)** The Dice Similarity Coefficient (DSC) between 4D CT and SPECT percentile ventilation is shown in box plot format for five independent percentile mask ranges. For the set of cases evaluated in this study, the highest correlation between SPECT ventilation and each of the 4D CT-derived methods lies within the lowest percentile range (1-20)%. **b)** DSC is shown for all cases, quantified only over the (1-20)% percentile mask region. Black arrows indicate patient cases with clinically recorded functional deficits, as determined from the nuclear medicine exams.

Also of note in figure 3.9 are the relatively low magnitudes for the set of DSC values among all 4DCT ventilation methods when measured in relation to the corresponding percentile SPECT ventilation distributions. For the set of cases included in this study, all calculated DSC values were  $< 0.7$ . However, closer inspection of the SPECT image data reveals that the relatively low correlation is most likely, at least in part, the result of airway deposition of the  $^{99m}Tc$  DTPA aerosol, a known artifact [2] that results in hot spot regions in the nuclear medicine images that are not physiologically representative of the underlying local lung function. A potential impact of this phenomenon is to effectively mask local cold spot regions of hypo-ventilation that would otherwise be present in the nuclear medicine images, but instead are represented as hot

spots that could potentially be mapped into the highest percentile mask regions. In those instances, the ability of the 4DCT-based imaging methods to identify such regions as belonging to the lowest functioning percentile masks results in poor correlation between the SPECT- and 4DCT-derived percentile distributions. In addition, the presence of large image intensities introduced by airway deposition of the  $^{99m}\text{Tc}$  DTPA aerosol effectively shifts the mapping of image intensities away from their true position by a magnitude that is dependent on the spatial extent and intensity magnitude of the airway deposition hot spots.



**Figure 3.10. CT airway and SPECT ventilation.** a) Two coronal CT sections are shown with the airway segmented in red (see arrows), and the corresponding  $^{99m}\text{Tc}$ -DTPA aerosol SPECT-CT overlain. b) A volumetric rendering of the segmented airway tree is also shown. c) Histograms are shown representing the distribution of CT-derived airway voxels versus percentile ventilation values for CT density-based ventilation and  $^{99m}\text{Tc}$ -DTPA aerosol SPECT ventilation. The figure

shows that airway voxels have the highest probability to be found within the  $\geq 90^{\text{th}}$  percentile SPECT ventilation. This finding supports our assertions that the  $^{99\text{m}}\text{Tc}$ -DTPA aerosol SPECT ventilation hot spots are due primarily to airway deposition of the radiopharmaceutical agent.

In light of these considerations, an additional analysis was performed in order to investigate the potential for airway deposition of aerosol particles to contribute to the SPECT image intensities within the lung parenchyma ROIs used to mask the 4DCT-based ventilation calculations. For each patient dataset included in this study, the CT components of the nuclear medicine SPECT-CT exams were segmented to remove the major airway tree by three-dimensional region growing from manually selected seed point locations placed within the trachea, above the level of the carina, as described above for airway tree removal from the ventilation lung parenchyma masks. These SPECT-CT airway trees were aligned with the maximum exhalation component image from each 4DCT set, by virtue of the CT-to-CT affine registration described in section 3.2.2. For both  $\text{sVent}_{\text{HU}}$  and  $^{99\text{m}}\text{Tc}$  DTPA aerosol SPECT percentile ventilation images, all percentile values that were common to the intersection of the SPECT-CT airway tree and 4DCT ventilation lung parenchyma masks were obtained and grouped for all cases into 10 equally spaced percentile bins. A set of histograms was then determined showing the distribution of percentile ventilation values for both images located within the SPECT-CT airway tree, and included in the ventilation analysis. Figure 3.10 shows the results for the seven patient cases.

In figure 3.10a, the registered SPECT ventilation image is shown in two coronal sections overlain the corresponding maximum exhalation component phase image. Also shown are multiple segments of the SPECT-CT airway tree object that overlap the same coronal slices. Arrows indicate regional hot spots in the nuclear medicine image, which correspond anatomically with the SPECT-CT airway structure. A volumetric rendering of the airway tree corresponding to the example case is shown in figure 3.10b (note that the rendering is not to the same scale as the anatomy shown in figure 3.10a). Figure 3.10c shows the histograms of percentile ventilation values for  $\text{sVent}_{\text{HU}}$  and  $^{99\text{m}}\text{Tc}$  DTPA aerosol SPECT ventilation that overlap with the SPECT-CT airway trees for all cases. For the nuclear medicine ventilation images, the histograms indicate that SPECT-CT airway voxels that are common to the ventilation lung parenchyma masks have the highest probability to be found within the  $\geq 90^{\text{th}}$  percentile SPECT ventilation. This finding supports our assertion that the  $^{99\text{m}}\text{Tc}$  DTPA aerosol SPECT ventilation hotspots

are due to airway deposition of the  $^{99m}\text{Tc}$  DTPA aerosol. In contrast, the relatively flat histogram values for  $\text{sVent}_{\text{HU}}$  suggest that all percentile ventilation values are equally likely to overlap the SPECT-CT airway voxels.

### 3.4 Discussion

In this study we compared ventilation images generated from 4DCT by three methods, a CT value method and two numerical implementations of the Jacobian method, with registered  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation images. We found both Jacobian- and density change-based measurements of specific ventilation obtained from the 4DLTM spatial transformation functions yielded good overall agreement with image segmentation-based measurements of the resting tidal volume. Qualitatively, however, we found subjective differences between the visual appearances of Jacobian- relative to density based-methods. The local subjective differences were quantified in terms of the spatial overlap of the lowest and highest functioning percentile mask regions among all combinations of the 4DCT-derived ventilation calculations. The conversion to percentile images for purposes of comparative evaluation among methods is particularly relevant in this context, as previous radiotherapy treatment planning studies have demonstrated the use of percentile representations of functional image information during inverse planning for functional lung avoidance (for example, see [16, 94, 97]). Using the Dice similarity coefficient as the spatial overlap metric, we found the Jacobian-based implementations to be highly correlated in both percentile mask regions, while correlation with  $\text{sVent}_{\text{HU}}$  was clearly less (see figure 3.8). Additionally, the spatial overlap was also assessed between each of the 4DCT ventilation methods with the clinical reference  $^{99m}\text{Tc}$  DTPA aerosol SPECT. Similar evaluation showed that the highest correlation among methods was in the lowest functioning lung regions, defined by the (0-20)% percentile masks. Furthermore, it was shown that  $\text{sVent}_{\text{HU}}$  yielded the highest correlation with the nuclear medicine exams, with the difference reaching statistical significance ( $p < 10^{-4}$ ).

The use of clinically acquired  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation exams as a reference for evaluation of 4DCT-based ventilation is complicated by the inherent qualitative nature of the nuclear medicine studies, as well as the potential for image artifacts resulting from airway deposition of the  $^{99m}\text{Tc}$  DTPA aerosol. In order to demonstrate the presence of airway deposition artifacts in our patient data set, volumetric airway tree segmentations were obtained from the CT component of each SPECT-CT study. Histograms were calculated showing the corresponding percentile ventilation



distribution within the airway tree volume to assess the likelihood of values to be found within that region. Figure 3.10c illustrates that the highest percentile SPECT ventilation is most likely to be found within the airway tree. In contrast, the spatial overlap of the airway tree was shown to have negligible impact on the calculated  $sVent_{HU}$ . An alternative validation strategy could potentially utilize clinically acquired SPECT ventilation exams as reference; though the use of a radiopharmaceutical agent that is less susceptible to airway deposition artifacts is highly recommended. DTPA is highly soluble in water, leading to its deposition in the airway wall.  $^{99m}Tc$ -Technegas is poorly soluble in water, and thus could potentially be used for this purpose, though regulatory approval in the United States is currently lacking, and the presence and degree of artifacts yet to be characterized.

Not surprisingly, the Jacobian-based implementations of specific volume behave in a very similar fashion. An examination of the relationship between the two approaches reveals that they are essentially the same method in that the only difference lies in the numerical approximations utilized for the partial derivatives of  $T$  when calculating the quantity  $J(\bar{x}_{inhale})$ . Put another way, the  $sVent_{GJ}$  method is simply a particular algorithmic instantiation of the  $sVent_{AJ}$  method. Inaccuracies in the deformable image registration immediately manifest themselves within both implementations as inaccurate approximations for the partial derivatives of  $T$  leading directly to errors in the specific volume calculation. The effect of DIR errors on the spatial accuracy of either ventilation method is a current research endeavor. Because one cannot assume a correlation between spatial mis-registration, and the corresponding error introduced by the intensities of the mis-registered voxel pair [99], it is not immediately obvious how spatial errors in the DIR will translate into degradation of the  $sVent_{HU}$ . Careful consideration of these effects, as well comparative evaluation among 4DCT-based ventilation images as a function of systematic DIR perturbation could potentially lead to the formulation of a clinical DIR QA strategy in which concordance of the spatial distribution of specific ventilation among 4DCT-based methods serves as a surrogate for relative DIR spatial accuracy performance. However, characterization of the effects of mis-registration on the quantitative specific ventilation should first be assessed for a larger patient cohort, with corresponding evaluation of the DIR spatial accuracy performance. The spatial accuracy of the 4DLTM DIR formulation utilized in this study was evaluated rigorously using large samples of manually selected feature pairs between 4DCT component images [98].

Fuld et al. define specific ventilation as the change in volume normalized by the initial air volume [104], given by:

$$sVol = \frac{\Delta V}{V_{ae}}, \quad (28)$$

where  $sVol$  is the specific ventilation,  $\Delta V$  is the volume change, and  $V_{ae}$  is the volume of air in the initial exhale phase. We utilize this definition of specific ventilation and have generalized their derivation to the discretized image case, where one voxel on exhale may map into  $N$  voxel on inhale, in the Appendix. We use this same definition for specific ventilation in our derivation of the Jacobian implementations described in sections 3.2.6 and 3.2.7. In contrast, Reinhardt et al. calculate the regional specific volume change from the Jacobian of the displacement fields [7], given by:

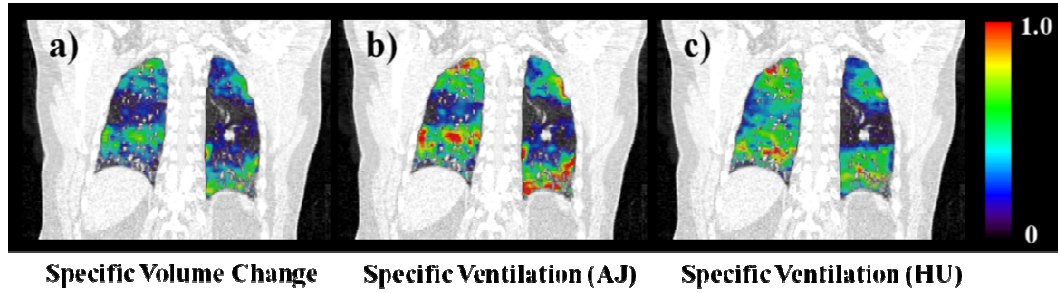
$$V_1 = J(x, y, z)V_0, \quad (29)$$

where  $V_0$  is the volume at position  $(x, y, z)$  in the initial image  $I_0$ ,  $V_1$  is the volume at the corresponding point in  $I_1$ , and  $J(x, y, z)$  is the Jacobian of the vector displacement field at point  $(x, y, z)$ . Kabus et al. [105] utilize a similar equation to represent the local volume change at position  $(x, y, z)$ . Our derivation of the Jacobian-based specific ventilation (see section 3.2.6) differs with regard to the reference volume. Both Reinhardt et al. and Kabus et al. reference the specific volume change to the initial voxel volume, which contains both lung tissue and air. In our derivation, we correct for the volume of air displaced by lung tissue; where the relation between the voxel volume and corrected initial air volume is given by:

$$V_{air} = \left( -\frac{HU}{1000} \right) V_{voxel}, \quad (30)$$

where  $V_{air}$  is the volume of air in the voxel located at position  $(x, y, z)$ ,  $V_{voxel}$  is the voxel volume, and  $HU$  is the CT value in Hounsfield Units at the point  $(x, y, z)$ . To convert

the specific volume change described by Reinhardt et al. and Kabus et al. to the specific ventilation, as derived in section 3.2.6, requires dividing each voxel by the factor in parenthesis in the above equation. Without this correction, the specific volume change images appear different from the CT-density derived specific ventilation images (see figure 3.11). We speculate that incorporating this change may improve the correlation of the Jacobian-based CT ventilation with the Xe-CT ventilation values found in Reinhardt et al. [7] and the correlation with segmentation-based volume changes found in Kabus et al. [106].



**Figure 3.11. Specific volume change versus specific ventilation.** a) The specific volume change was calculated using the analytic implementation of the Jacobian operator (equation (21)) and is shown overlain the corresponding coronal segment from the maximum exhalation phase 4D CT image. b) The specific ventilation  $sVent_{AJ}$  given in equation (22) is shown overlain the same coronal section for comparison. c) The  $sVent_{HU}$ , given in equation (16) is also shown. Note that all images are quantitative and presented on the same intensity scale (shown on right).

There is much interest to utilize ventilation images in radiotherapy treatment planning of thoracic malignancies [16, 95, 97, 107-109]. In a prospective study, Abratt et al. demonstrated delivery of radiotherapy through hypo-perfused pulmonary regions for lung cancer treatment results in less pulmonary toxicity [1]. In that study  $^{99m}Tc$ -macro-aggregated albumin ( $^{99m}Tc$ -MAA) injected intravenously was the imaging agent and planar scintigraphy the image acquisition method.  $^{99m}Tc$ -MAA provides nearly artifact free images of pulmonary perfusion and has been utilized extensively to study radiation lung injury [45, 108, 109]. Perfusion and ventilation defects in lung cancer arise due to tumor compression of blood vessels, obstruction of the airway, or physiological response producing one defect type in response to the other. Lung cancer does not appear instantly, primary tumors grow on the time frame of many months to years. As long-standing

lesions develop, they are likely to produce defects in both perfusion and ventilation. Narabayshi et al. [110] found 80% matched ventilation and perfusion defects in 18 lung cancer patients studied with SPECT imaging. Suga et al. [111] evaluated respiratory gated  $^{99m}\text{Tc}$ -MAA SPECT perfusion and gated  $^{99m}\text{Tc}$ -Technegas SPECT ventilation in 23 patients with lung cancer. Each patient was found to have both a ventilation and perfusion defect associated with their primary tumor. In this study we compare methods to obtain ventilation images from no contrast added 4DCT images, the type widely used for radiotherapy treatment planning. We anticipate the findings of Abratt et al. with perfusion imaging will also hold with ventilation imaging due to the concordance of defects. However, this hypothesis that ventilation images in treatment planning will reduce clinical toxicity in treated patients remains to be tested. The 4DCT ventilation images described in this study, derived from the treatment-planning 4DCT images, require no additional imaging sessions or irradiation of the patients and are available simply for the cost of an additional computational step. In addition, these images derived from CT images acquired for treatment planning do not require multi-modality image registration, a potential source of error. Using these images, the functional heterogeneity of the lungs can be taken into account, producing radiotherapy treatment plans which may reduce normal tissue complications. Reducing clinical toxicity with ventilation images remains to be demonstrated.

### 3.5 Conclusion

In this study we compared quantitative images of pulmonary specific ventilation generated from 4DCT by three methods, a CT value method and two implementations of the Jacobian method, with registered  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation images. All 4DCT-based images of specific ventilation showed good linear correlation with image segmentation-based measures of the global tidal volume. Moreover, the density change-based method showed significantly greater spatial correlation with the clinical reference nuclear medicine exams. In addition, we demonstrated experimentally that the  $s\text{Vent}_{\text{GJ}}$  is a special instance of the general  $s\text{Vent}_{\text{AJ}}$  formulation, differing only in the numerical implementation. Future work will directly address the impact of spatial registration errors in the DIR on degradation of both density change- and Jacobian-based ventilation methods.] " [112]

## Chapter 4

### Spatial Correlation of 4D CT Ventilation and SPECT Pulmonary Perfusion Defects in Patients with Malignant Airway Stenosis

#### 4.1 Introduction

Physiologic validation of 4D CT ventilation with the current clinical standard nuclear medicine single-photon-emission computed tomography (SPECT) ventilation imaging is necessary. In initial works by Guerrero et al. [14, 15], preliminary evaluation was based on the correlation of tidal volume calculations derived from the ventilation images with corresponding measurements obtained by manual segmentation of lung voxels from the CT image pairs. In both studies, linear correlation between measured and mass-corrected tidal volume calculations was demonstrated, with corresponding  $R^2 = 0.97$  for comparison over 22 (breath-hold CT) and 3 (4DCT) clinically acquired patient datasets, respectively.

Castillo et al. further extended the ventilation framework presented by Guerrero et al. [14-16], providing explicit mathematical derivation of an expression for the local specific ventilation within any exhalation state lung voxel in terms of registered HU between corresponding inhalation and exhalation breathing states [112]. In that study, a second methodology was also investigated, similar to the formulation originally proposed by Reinhardt et al. [7, 113] in which regional expansion and contraction of lung volume elements is determined directly from the Jacobian of the calculated DIR displacement field linking the respiratory phase images. Both methods were evaluated for 7 patient data sets, in which the percentile distributions of ventilation were compared with those obtained from  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation scans acquired on the same day as the 4DCTs used for ventilation quantification. Using the Dice similarity coefficient as the spatial overlap metric, the study found significantly ( $p < 10^{-4}$ ) higher correlation between HU-based derivation of specific ventilation with the clinical reference SPECT. Furthermore, the results demonstrated that the highest correlation with SPECT ventilation was achieved in the lowest functioning pulmonary regions (as defined on SPECT) for cases in which a marked functional defect due to airway obstruction was noted in the clinical record. These findings suggest an IGRT strategy for incorporation of the CT lung function images into a treatment planning framework for lung cancer patients, in which hypo-functioning regions are identified in order to remove avoidance restriction (dysfunctional allowance). This strategy could be combined with function-based conformal avoidance methods previously described [16, 97, 114, 115].

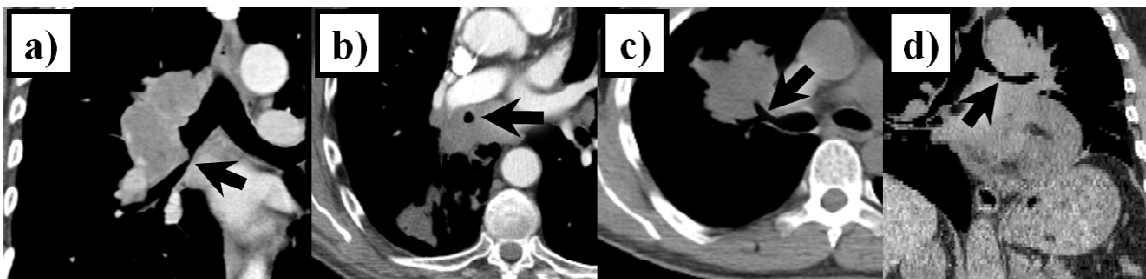
It is the focus of the present work to extend the work of Castillo et al. [112] in order to assess the spatial correlation of hypo-functioning defect regions identified on 4DCT ventilation

with clinically acquired nuclear medicine SPECT perfusion in a patient population with malignant airway stenosis. We hypothesize that hypo-ventilated regions will correlate spatially with hypo-perfused pulmonary regions distal to the known obstructing lesions. To test this hypothesis, 10 patient datasets from patients with malignant airway stenosis are examined, for whom 4DCT and perfusion SPECT images have been obtained prior to the initiation of thoracic radiotherapy. Using SPECT perfusion as the reference, the spatial overlap of corresponding defect regions will be assessed using the Dice similarity index [102].

## 4.2 Materials and Methods

### 4.2.1 Thoracic 4D CT Images

The set of patient images utilized in this study was acquired as part of a 4DCT ventilation validation imaging study, with approval by the M. D. Anderson Institutional Review Board (protocol 2006-0698). Patients were selected retrospectively and were limited to include lung cancer patients with radiographically demonstrated airway obstruction (figure 4.1). All study imaging was performed prior to the initiation of treatment. The 4DCT datasets were acquired at 2.5 mm slice spacing using a General Electric Discovery ST PET/CT scanner (GE Medical Systems, Waukesha, WI). All 4DCT images were acquired to include the entire thorax with voxel dimensions  $(0.97 \times 0.97 \times 2.50) \text{ mm}^3$ . The 4DCT cine acquisition technique using a respiratory signal from the Real-Time Position Management Respiratory Gating System (Varian Medical Systems, Palo Alto, CA) has been previously described [93]. Clinical diagnoses for the 10 cases utilized in this study are given in table 4.1.



**Figure 4.1. Malignant airway stenosis.** Patients were selected retrospectively, and were limited to include lung cancer patients with radiographically demonstrated airway obstruction. In the examples shown, narrowing is indicated by an arrow in the **a)** bronchus intermedius, **b)** right lower lobe bronchus, **c)** distal to the right upper lobe bronchial bifurcation, and **d)** left main bronchus.

#### 4.2.2 Pulmonary Perfusion SPECT Imaging

Pulmonary perfusion SPECT-CT images were acquired using a Siemens Simbia T6 SPECT-CT scanner (Siemens, Hoffman Estates, IL) following intravenous administration of approximately 5 mCi of Technetium-99m-labeled macro-aggregated albumin ( $^{99m}\text{Tc}$  MAA), in the same position as that of the CT ventilation study. The  $^{99m}\text{Tc}$ -labeled MAA SPECT acquisitions were performed with the patients in the supine position with normal resting breathing. Each acquisition covered the entire lung volume in the emission image. Co-registered CT images were also acquired to perform attenuation correction of the SPECT images, and for affine registration with the maximum exhalation component phase image from the corresponding 4DCT planning data set, using the CT-to-CT fusion software in a commercial radiotherapy treatment planning system (Pinnacle, version 8.1x, Philips Medical Systems, Andover, MA, USA). The affine coefficient matrix for each case was subsequently used to align the SPECT perfusion and 4DCT ventilation function images for spatial overlap analysis. The use of affine registration for functional and anatomical alignment has been previously reported in radiotherapy treatment planning studies requiring multi-modality image fusion or spatial registration of anatomical and quantitative functional image information (see, for example [16, 94, 95, 97, 112]).

#### 4.2.3 MILO Deformable Image Registration

Recently we reported on a novel DIR algorithm referred to as Moving Least Squares Guided Local Optimization (MILO) that was designed specifically to account for difficulties associated with deformable registration of breath-hold CT image pairs acquired as part of the National Heart Lung Blood Institute (NHLBI) COPDgene study (<http://www.copdgene.org>) [115]. Deformable registration of the COPDgene breath-hold CT image pairs is particularly challenging relative to 4D CT images due to the relatively large displacements, change in density and CT value, the difference in image noise, the highly non-uniform mechanical properties of lung tissue in patients with COPD, and the changes in anatomic shape of the vasculature due to the large volume change. The MILO algorithm is formulated based on the general assumption that the change in intensity between corresponding voxels in the reference and target image is additive and unknown. The algorithm is then based on iteratively solving a sequence of small, local, nonlinear least squares problems defined on a subset of all voxels contained in the image domain. The solution to the sub-problem minimizes the variance of the local image intensity mismatch and describes the motion of a single volume element. At each iteration, the initial guess for each local sub-problem is provided by a moving least squares (MLS) estimate. The full DIR solution is then the displacement field represented by the MLS interpolant of the sub-problem solutions.

Though it was not designed specifically within the context of DIR for 4D CT images, we anticipate high spatial accuracy performance of the MILO algorithm over these data sets, due to the inherent set of computational challenges it was designed to overcome. In this study, we perform preliminary evaluation of the MILO spatial accuracy performance using the reference 4D CT datasets described in sections 2.2.3 and 2.2.4.

#### 4.2.4 Lung Parenchyma Segmentation

Initially, CT values  $\leq -250$  HU were selected as representing lung parenchyma voxels. The trachea and main stem bronchi were separately segmented by three-dimensional morphological growing from manually placed seed positions, and subsequently removed [63]. A manually placed seed point within the heart was similarly grown to extract pulmonary vasculature structures. For all cases, the primary lung tumor volume was also excluded from the final lung masks. The measured volumes of the segmented lungs for each case are later used as reference with which to compare corresponding tidal volume calculations derived from the quantitative function images.

#### 4.2.5 Pulmonary Ventilation from 4D CT

In 2000, Simon presented a general methodology for obtaining pulmonary function values, such as the specific compliance (compliance per unit volume), from pairs of breath-hold CT images, based on the correspondence of average CT values in manually registered parenchymal sub-volumes [13]. Based on the simple principle that the lung can be modeled as a linear combination of air and tissue compartments, Simon demonstrated that certain pulmonary physiologic mechanisms could be quantified regionally by exploiting their impact on the local radiographic CT Hounsfield Unit (HU), in images acquired at distinct physiologic states (e.g., inhalation and exhalation breath-hold). Previous work extended Simon's framework to provide explicit quantification of regional specific ventilation within any exhalation state lung voxel in terms of registered HU between corresponding inhalation and exhalation breathing states [112]:

$$\frac{\Delta V}{V_{exhale}^{air}} = 1000 \frac{(\bar{H}_{inhale} - HU_{exhale})}{HU_{exhale} (1000 + \bar{H}_{inhale})}, \quad (31)$$



where  $\bar{H}_{inhale}$  represents the mean HU of tissue volume that is mapped into the exhale voxel with corresponding CT number given by  $HU_{exhale}$ . The quantity  $(\Delta V/V_{exhale}^{air})$  is the fractional volume change due to inspiration, i.e., the local specific ventilation.

For the images included in this study, a volumetric DIR solution was determined between the extreme component phases from each 4DCT, where the 50% phase was considered maximum exhalation, and the 00% phase considered maximum inhalation during resting tidal breathing. The spatial correspondence of tissue elements between the different breathing states was defined such that for each parenchymal volume element  $\bar{x}_{inhale}$  in the inhalation state image, the equivalent coordinate position in the exhalation state is given by:

$$\bar{x}_{exhale} \equiv \Phi(\bar{x}_{inhale}) = \bar{x}_{inhale} + T(\bar{x}_{inhale}), \quad (32)$$

where  $T$  is the vector displacement field associated with the spatial transformation function  $\Phi$ .

Given the solution  $\Phi$ , equation (31) is then applied at each exhale voxel position defined by the lung parenchyma mask. The result is a parametric map, co-registered with the exhalation CT, in which voxel intensities represent fractional change in air content within the originating compressed sub-volume that occurred due to inspiration. As described in section 3.2.5, a correction factor (equation (17)) is first applied to each lung voxel in the inspiration state image to account for the observed difference in CT-derived mass between the two images. Local  $3 \times 3 \times 3$  voxel averages were used for each HU input into equation (31) to account for noise in the CT image intensities, as well as small spatial errors in the DIR. The final three-dimensional ventilation images were created following smoothing with a  $9 \times 9 \times 3$  voxel box average filter.

Next, percentile images were generated to facilitate inter-modality spatial overlap analysis. Both ventilation and perfusion function images were initially mapped into percentile distribution images in which each voxel value was replaced by the value given by the corresponding cumulative distribution function (CDF), scaled to the interval  $[0, 100]$ . For observed image intensity  $i$ , the corresponding  $CDF(i)$  is given by:

$$CDF(i) = \frac{N_i \in [0, i_{\max}]}{N_{total}}, \quad (33)$$

where  $i_{\max}$  is the maximum image intensity and  $N_{total}$  is the total number of voxels within the segmented image mask. At each position  $\bar{x}$ , the percentile image  $P(\bar{x})$  is scaled to the interval  $[0, 100]$ , according to:

$$P(\bar{x}) = 100 \times CDF(F(\bar{x})), \quad (34)$$

where  $F(\bar{x})$  is the image intensity at position  $\bar{x}$  within the corresponding lung function image (i.e., ventilation or perfusion). For the SPECT perfusion datasets, percentile images were derived only from the set of image intensities included in the lung parenchyma segmentations used to mask each of the 4DCT ventilation calculations [112].

#### 4.2.6 Spatial Overlap of VQ Functional Defects

A semi-automated segmentation scheme was used to identify contiguous regions of pulmonary hypo-perfusion resulting from airway obstruction by the gross tumor volume. For each case, the 50% 4DCT phase image, as well as the raw and percentile SPECT perfusion images were imported into a research version of the Pinnacle treatment planning system (version 8.1x, Philips Medical Systems, Andover, MA, USA). The auto-segmentation feature of this system was used to define a percentile threshold for each case, sufficient to encompass the visible defect region. Using both the raw and percentile perfusion images for guidance, manual trimming of the segmented percentile volume was subsequently performed to include only the specific defect region distal to the known airway malignancy. This process ensured the removal of artificial cold spots present in the raw perfusion images, resulting from small spatial alignment errors during the affine registration process. For each case, region of interest coordinate information was subsequently exported and reconstructed into a binary image volume for spatial overlap analysis.

Semi-automated segmentation of regional pulmonary hypo-ventilation was similarly performed. For each case, the 50% 4DCT phase image, as well as the raw and percentile 4DCT ventilation images were imported into the Pinnacle TPS. Initially, the percentile threshold defined previously for segmentation of the corresponding perfusion defects was applied to the percentile ventilation images. Three independent observers then performed the manual trimming process, to include only ventilation defects distal to the known airway malignancy. The manual modifications were performed independently by each observer, and without prior knowledge of the SPECT perfusion segmentation regions, save for the initial percentile threshold masking. Each of the three region of interest files was subsequently exported and reconstructed for spatial

overlap analysis with the reference SPECT, as well as assessment of the corresponding inter-observer uncertainty.

The quantitative Dice similarity coefficient (DSC) between the reference SPECT defect mask and each of the three corresponding ventilation masks was determined for all cases. The DSC is a measure of the degree of overlap between a reference, and delineated area or volume and is defined as the ratio of twice the volume of intersection to the sum of the two volumes [102, 103]. For “reference” ( $V_{\text{Reference}}$ ) and “delineated” ( $V_{\text{delineated}}$ ) volumes, the DSC is given by:

$$DSC(V_{\text{Reference}}, V_{\text{delineated}}) = \frac{2 \times |V_{\text{Reference}} \cap V_{\text{delineated}}|}{|V_{\text{Reference}}| + |V_{\text{delineated}}|}. \quad (35)$$

To account for sensitivity of the DSC to small changes in volume when the volumes being compared are small, as well as uncertainty in manual segmentation of the reference mask (in this case, the SPECT perfusion defect masks), previous investigators have utilized the normalized Dice similarity coefficient (NDSC), defined as the DSC of the reference volume with the reference volume contracted by 1 mm ( $\text{NDSC}_{1\text{mm}}$ ), or the width of a CT pixel in the transverse plane [103, 116]. The normalized DSC provides reference for interpreting the measured spatial overlap in terms of a quantifiable spatial discrepancy. In the present study, mean and standard deviation DSC,  $\text{NDSC}_{1\text{mm}}$ , and  $\text{NDSC}_{2\text{mm}}$  are computed for each case to assess the overall spatial correspondence of regional pulmonary ventilation-perfusion defects due to malignant airway stenosis.

### 4.3 Results

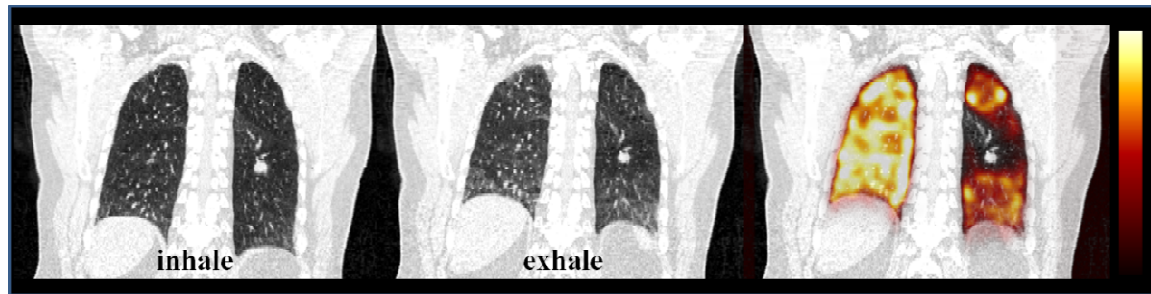
#### 4.3.1 4D CT and $^{99\text{m}}\text{Tc}$ MAA SPECT Image Properties

Table 4.1 provides clinical diagnoses for all cases included in this study, as well as volume and corresponding anatomic segment containing hypo-functioning perfusion defect regions. Figures 4.2a & b show coronal images from the extreme inspiration and expiration phases of a typical 4DCT image set (case #1 in this study). The corresponding  $^{99\text{m}}\text{Tc}$  MAA SPECT perfusion image is shown in figure 4.2c superimposed on the extreme expiration phase CT image, following affine registration as described in section 4.2.2.

| Case # | Diagnosis | Obstruction           | Defect Volume (mL) | Defect Volume (%) |
|--------|-----------|-----------------------|--------------------|-------------------|
| 1      | SCLC      | LLL; Superior Segment | 157.3              | 22.3              |
| 2      | NSCLC     | Right Lung            | 372.9              | 87.3              |
| 3      | NSCLC     | RLL                   | 335.2              | 27.6              |
| 4      | NSCLC     | Left Lung             | 629.3              | 74.1              |
| 5      | NSCLC     | RUL / RML             | 448.2              | 35.1              |
| 6      | NSCLC     | RLL; Superior Segment | 77.3               | 5.7               |
| 7      | NSCLC     | RUL; Superior Segment | 191.9              | 18.3              |
| 8      | NSCLC     | RLL                   | 192.1              | 20.7              |
| 9      | NSCLC     | Left Lung             | 576.7              | 93.7              |
| 10     | NSCLC     | RUL                   | 329.7              | 37.8              |

abbreviations: SCLC = small cell lung cancer; NSCLC = non-small-cell lung cancer; LLL = left lower lobe; RUL = right upper lobe; RML = right middle lobe; RLL = right lower lobe

**Table 4.1. Clinical diagnosis and perfusion defect characteristics.** Clinical diagnoses are shown for each case included in this study. The anatomic region distal to the tumor obstruction is listed. Corresponding volumes of the contoured perfusion defects are also shown, both in units of milliliters as well as in percentage of the total segmented volume of the affected lung.



**Figure 4.2. 4D CT and SPECT perfusion.** a) A maximum inspiration phase 4D CT image is shown in coronal section. b) The corresponding section is shown for the maximum expiration phase 4D CT image. c) The  $^{99m}\text{Tc}$  MAA pulmonary perfusion SPECT-CT was registered to the maximum expiration phase CT image using affine registration. The corresponding coronal section is shown superimposed, with color scale indicated. The SPECT perfusion demonstrates a functional defect within the superior segment of the left lower lobe.

The lung volumes, tidal volume, and average CT value in HU were determined from the segmented CT images, with measured values reported in table 4.2. The MILO DIR was applied to link the images from the extreme inhale to the extreme exhale phases. The average (and standard deviation) lung voxel displacements were obtained directly from the calculated MILO transformation functions and are also reported in table 4.2. Tidal volumes ranged from 224 to 953

mL, while the average calculated lung voxel displacement ranged from 4.94 to 12.70 mm. The variability in the mechanical and physiologic lung properties depicted in table 4.2 reflects the inherent inter-subject variability in lung function characteristics, as well as in the extent and severity of the underlying malignancy. It is noteworthy that cases #4 and #6 measure nearly identical average HU within each of the inhale and exhale lung volumes, despite widely different measurements of the resting tidal volume. While, in general, we do not expect a strict relationship between absolute change in CT number and tidal volume, nearly identical CT numbers in both images suggest similar volume change. The explanation for the observation in this case is twofold. The presence of 4D CT reconstruction artifacts in the case #6 T00 phase image, in which multiple misplaced cine segments yielded artificially high CT numbers within the segmented inhale lung, ultimately resulted in an average inhale HU that was artificially high. In addition, the same artifacts resulted in repetitive cine representation of thoracic structures such as the heart, diaphragm, and liver, thereby reducing the segmented lung volume. These two effects of the image acquisition yielded artificially high inhale HU and artificially low measured tidal volume. A representative CT slice from the case #6 T00 phase image is shown in sagittal section in figure 4.9 alongside an illustration of the subsequent impact on the resulting ventilation map.

| Case # | Inhale Vol (ml) | Exhale Vol (ml) | Tidal Vol (ml) | Avg (SD) Displ (mm) | Inhale Avg (HU) | Exhale Avg (HU) |
|--------|-----------------|-----------------|----------------|---------------------|-----------------|-----------------|
| 1      | 2028            | 1599            | 429            | 9.54 (5.10)         | -747            | -676            |
| 2      | 1484            | 1260            | 224            | 6.80 (2.81)         | -599            | -547            |
| 3      | 3122            | 2465            | 657            | 10.45 (6.70)        | -793            | -759            |
| 4      | 3285            | 2702            | 583            | 12.70 (11.07)       | -778            | -751            |
| 5      | 2874            | 2529            | 345            | 4.94 (3.85)         | -873            | -856            |
| 6      | 2929            | 2642            | 287            | 5.29 (2.81)         | -775            | -759            |
| 7      | 2810            | 1857            | 953            | 10.14 (4.35)        | -767            | -642            |
| 8      | 2176            | 1912            | 264            | 6.09 (2.92)         | -772            | -747            |
| 9      | 2767            | 2294            | 473            | 9.05 (3.33)         | -750            | -716            |
| 10     | 2252            | 1808            | 444            | 6.88 (4.69)         | -757            | -729            |

abbreviations: SD = standard deviation; HU = Hounsfield Unit

**Table 4.2. 4D CT characteristics.** Functional measurements obtained from the segmented lung mask regions are shown for all cases. Average displacements were determined from the calculated MILO DIR displacements within the lung masks.

#### 4.3.2 Deformable Image Registration Spatial Accuracy Assessment

The image and DIR reference landmark characteristics for the evaluation test cases utilized in this study are summarized in table 4.3. The data acquisition and statistical characterization of the

reference data for the thoracic 4D CT image sets has been described in detail (see sections 2.2.3 and 2.2.4).

| Case # | Image Dimension | Voxel Dimension | # Landmarks | Avg (SD) Displacement (mm) | Observer Error (mm) |
|--------|-----------------|-----------------|-------------|----------------------------|---------------------|
| ref-1  | 256×256×94      | 0.97×0.97×2.5   | 1280        | 4.01 (2.91)                | 0.85 (1.24)         |
| ref-2  | 256×256×112     | 1.16×1.16×2.5   | 1487        | 4.65 (4.09)                | 0.70 (0.99)         |
| ref-3  | 256×256×104     | 1.15×1.15×2.5   | 1561        | 6.73 (4.21)                | 0.77 (1.01)         |
| ref-4  | 256×256×99      | 1.13×1.13×2.5   | 1166        | 9.42 (4.81)                | 1.13 (1.27)         |
| ref-5  | 256×256×106     | 1.10×1.10×2.5   | 1268        | 7.10 (5.14)                | 0.92 (1.16)         |

**Table 4.3. CT image and reference data characteristics.** The image and voxel dimensions are shown for the set of DIR reference evaluation cases included in this study. Also shown are the number of spatial accuracy reference landmarks for each case, along with corresponding average (and standard deviation) landmark displacement. Estimates of observer variance in reference landmark registration were obtained by repeat registration as described in section 2.2.4, and are also shown as mean (and pooled standard deviation), combined for the set of multiple observers. All measurements of distance are reported in units of millimeters.

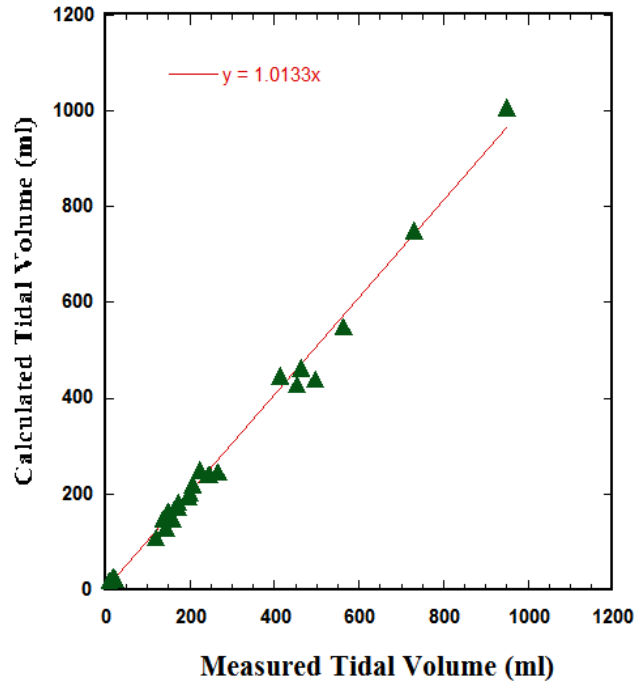
Table 4.4 shows a complete summary of the measured DIR spatial errors for the MILO algorithm over the set of five reference evaluation cases utilized in this study. Measured errors are shown in component RL, AP, and SI directions, as well as in three-dimensional Euclidean magnitude. Mean (and standard deviation) Euclidean errors ranged from 0.69 (0.96) – 1.26 (1.50) mm per case, while for the combined set of 4D CT test data ( $N = 6762$  reference measurements) mean spatial registration error was 0.92 (1.14) mm. The non-parametric Wilcoxon rank sum test was used to compare measured errors, matched between the MILO algorithm and repeated observer measurements to compare results between algorithm and observer. Table 4.4 shows that the observer errors measured by repeat registration on the 4D CT images are statistically equivalent to that of the MILO DIR ( $p > 0.05$ ), suggesting that the high spatial accuracy of the MILO algorithm has effectively achieved the error resolution of our measurement technique over this set of cases. These results reflect comparisons for each of the five subjects individually, and using all subjects simultaneously in the analysis. All results presented in table 4.4 were corrected for multiple comparison using the procedure of Hochberg [117].

| Case #                | Algorithm     | RL (mm)            | AP (mm)            | SI (mm)            | 3D (mm)            |
|-----------------------|---------------|--------------------|--------------------|--------------------|--------------------|
| ref-1                 | <i>MILO</i>   | 0.27 (0.47)        | 0.28 (0.46)        | 0.37 (0.90)        | 0.78 (0.96)*       |
|                       | <b>No DIR</b> | <b>0.58 (0.62)</b> | <b>0.67 (0.79)</b> | <b>3.68 (3.04)</b> | <b>4.01 (2.91)</b> |
| ref-2                 | <i>MILO</i>   | 0.27 (0.53)        | 0.26 (0.54)        | 0.27 (0.78)        | 0.69 (0.96)*       |
|                       | <b>No DIR</b> | <b>0.73 (0.85)</b> | <b>0.72 (0.88)</b> | <b>4.09 (4.37)</b> | <b>4.65 (4.09)</b> |
| ref-3                 | <i>MILO</i>   | 0.34 (0.57)        | 0.33 (0.57)        | 0.36 (0.89)        | 0.85 (1.03)*       |
|                       | <b>No DIR</b> | <b>1.17 (1.05)</b> | <b>1.28 (1.23)</b> | <b>6.10 (4.49)</b> | <b>6.73 (4.21)</b> |
| ref-4                 | <i>MILO</i>   | 0.46 (0.69)        | 0.52 (0.77)        | 0.60 (1.40)        | 1.26 (1.50)*       |
|                       | <b>No DIR</b> | <b>0.94 (1.21)</b> | <b>1.42 (1.22)</b> | <b>8.98 (5.04)</b> | <b>9.42 (4.81)</b> |
| ref-5                 | <i>MILO</i>   | 0.44 (0.65)        | 0.49 (0.73)        | 0.48 (1.07)        | 1.13 (1.22)*       |
|                       | <b>No DIR</b> | <b>0.86 (0.96)</b> | <b>1.74 (1.67)</b> | <b>6.30 (5.45)</b> | <b>7.10 (5.15)</b> |
| <b>4DCT-Composite</b> | <i>MILO</i>   | 0.35 (0.58)        | 0.37 (0.62)        | 0.41 (1.01)        | 0.92 (1.14)*       |
|                       | <b>No DIR</b> | <b>0.86 (0.95)</b> | <b>1.15 (1.19)</b> | <b>5.73 (4.53)</b> | <b>6.29 (4.28)</b> |

**Table 4.4. DIR spatial error summary.** Mean (and standard deviation) spatial registration errors are shown for the set of five 4D CT reference evaluation cases. Note that an asterisk (\*) indicates those cases for which the measured MILO spatial errors are statistically indistinguishable ( $p > 0.05$ ) from the estimates of observer variance obtained by repeat registration. The unregistered landmark distances (i.e., “No DIR”) are also shown for reference.

#### 4.3.3 Specific Ventilation from 4D CT

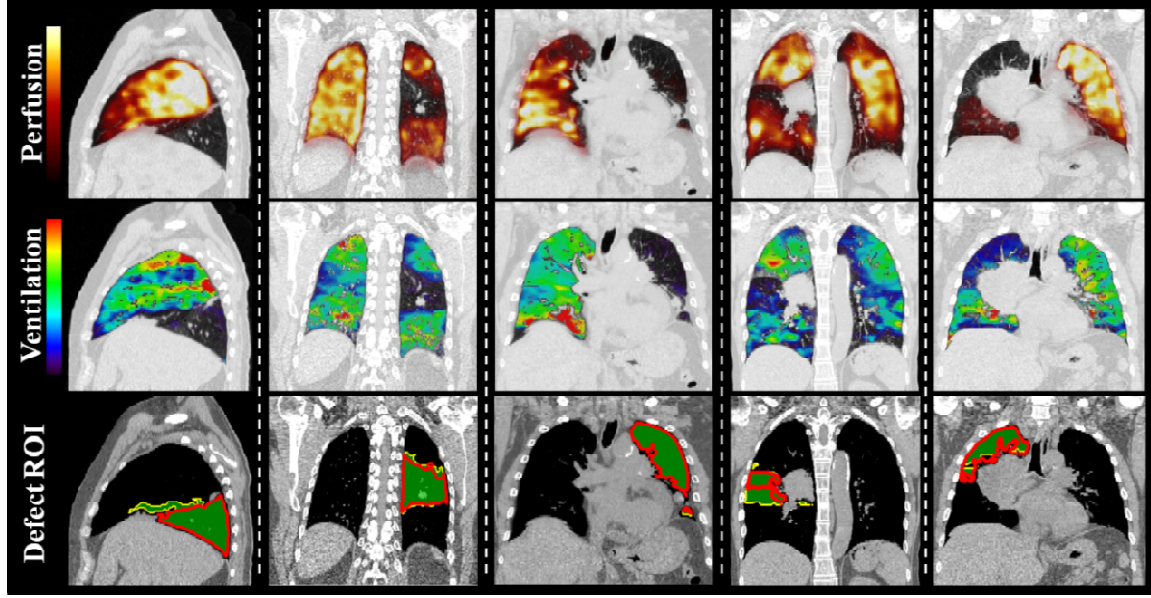
Figure 4.3 shows a comparison of calculated tidal volumes obtained directly from the specific ventilation images with measured values derived from the volumes of the binary lung parenchyma masks. Note that right and left lung masks were treated separately for each case, so as to provide more data points for self-consistency evaluation. Linear regression analysis resulted in a slope of 1.01 ( $R^2 = 0.99$ ) for the density-based measure of specific ventilation, indicating that on a global scale, the function images yielded good agreement with image segmentation-based measures of the resting tidal volume.



**Figure 4.3. Measured versus calculated tidal volumes.** Estimates of the resting tidal volume were obtained by manual segmentation of the lung parenchyma voxels, and subsequently compared to corresponding calculations derived from the 4D CT ventilation images. Linear regression analysis shows a best-fit to the data with slope equal to 1.01 ( $R^2 = 0.99$ ).

4DCT-derived images of the local specific ventilation are shown in figure 4.4 (middle panel) for a subset of five lung cancer patients included in this study with malignant airway disease causing reduced ventilation and perfusion within anatomic regions distal to their sites of obstruction. The sample cases shown are illustrative of macroscopic focal defect regions in the right upper lobe, right lower lobe, superior segment of the right lower lobe, superior segment of the left lower lobe, and left lung. The corresponding  $^{99m}\text{Tc}$  MAA SPECT perfusion images are also shown (top panel) for reference. Both ventilation and perfusion images depict spatially heterogeneous distribution of function in pulmonary regions not directly impacted by gross tumor involvement. The bottom panel of figure 4.4 shows the spatial correspondence of the segmented defect regions, overlain the exhalation phase image from each 4D CT. Visually, the defect ROIs appear highly correlated, suggesting the use of 4D CT ventilation for identifying hypo-functioning pulmonary regions in malignant airway disease.

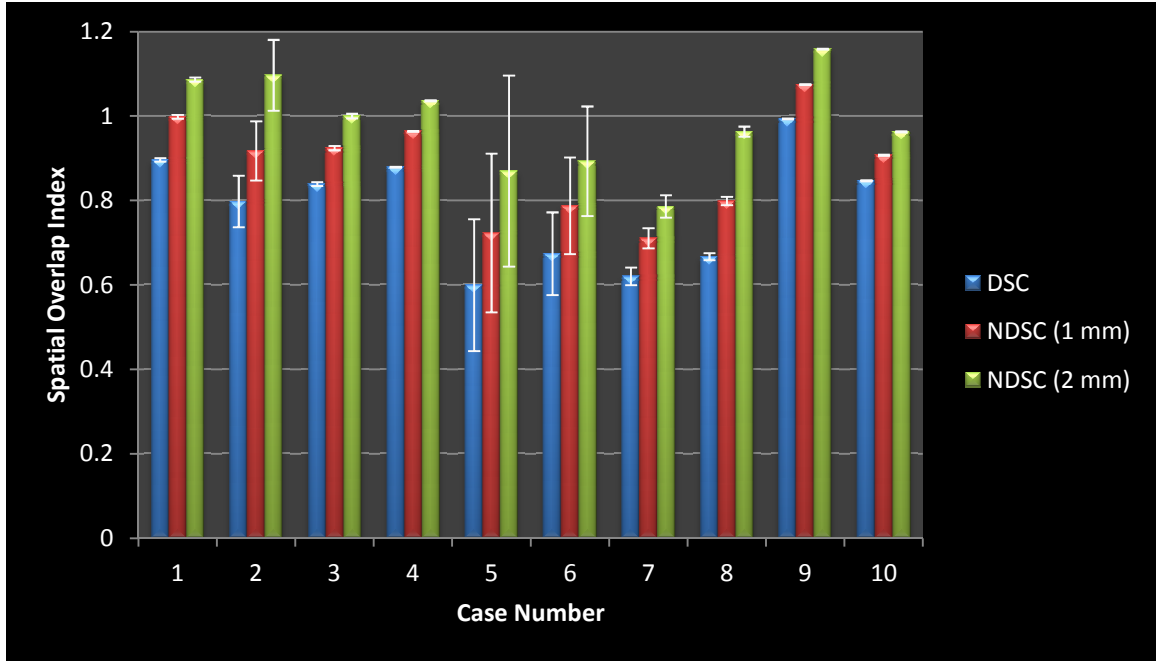




**Figure 4.4. Malignant airway disease causes hypo-functioning anatomic regions.**  $^{99m}\text{Tc}$  MAA SPECT perfusion and 4D CT ventilation images are shown for five cases with malignant airway disease. Regions of interest corresponding to the hypo-functioning regions (red = perfusion, green = ventilation) were constructed semi-automatically for each case and their overlap illustrated in the bottom panel. The hypo-functioning regions were **a)** the right lower lobe, **b)** the superior segment of the left lower lobe, **c)** the left lung, **d)** the superior segment of the right lower lobe, and **e)** the right upper lobe.

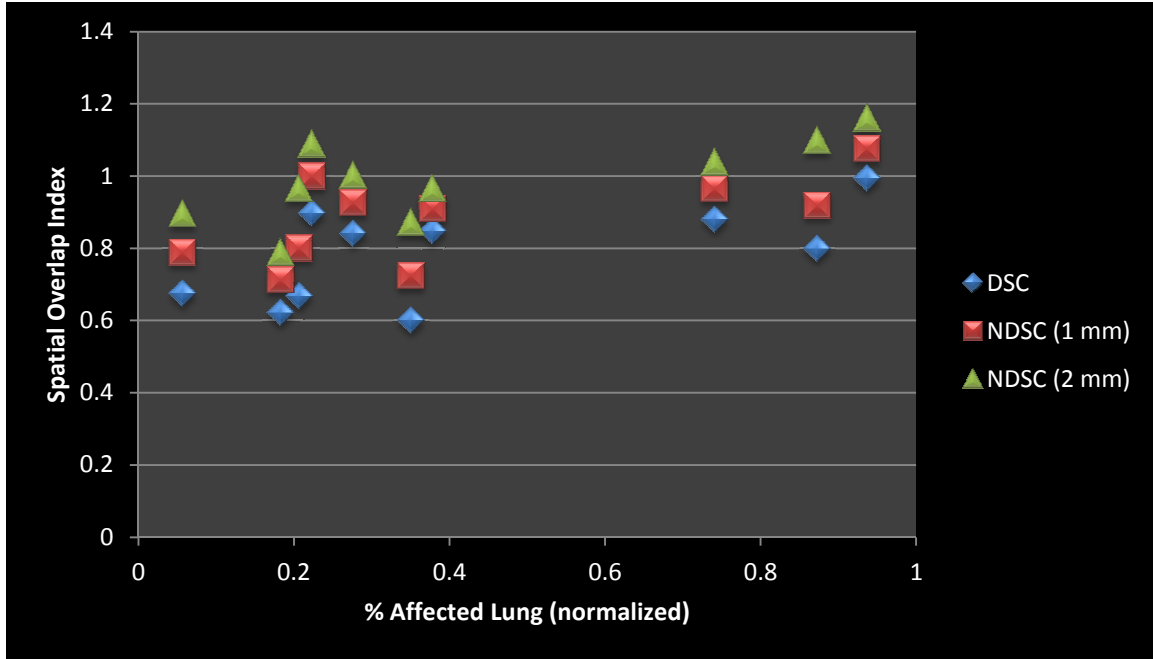
#### 4.3.4 Quantitative Spatial Overlap Comparison with $^{99m}\text{Tc}$ MAA SPECT Perfusion

Figure 4.5 shows a bar-graph representation of the average ( $\pm$  standard deviation) DSC,  $\text{NDSC}_{1\text{mm}}$ , and  $\text{NDSC}_{2\text{mm}}$  among observers for all cases included in this study. Average DSC values ranged from 0.60-0.99, with maximum coefficient of variation equal to 26% (range: 0.10-26%). Average normalized spatial overlap indices ranged from 0.71-1.07 and 0.79-1.16 for respective  $\text{NDSC}_{1\text{mm}}$  and  $\text{NDSC}_{2\text{mm}}$ , with corresponding coefficients of variation for all metrics similarly in the range: 0.10-26%. Note that  $\text{NDSC}_{1\text{mm}} > 1.0$  indicates the DSC of the contoured ventilation defect region exceeds that of the corresponding uncertainty index, implying that the ventilation and perfusion defect volumes agree to  $< 1$  mm uncertainty. Average DSC for the set of cases was 0.78, which demonstrates good spatial overlap agreement [118, 119]. The corresponding average  $\text{NDSC}_{1\text{mm}}$  was 0.88, while average  $\text{NDSC}_{2\text{mm}}$  was 0.99, indicating that on average, spatial overlap agreement between ventilation and perfusion defect regions was comparable to the threshold for agreement within 1-2 mm uncertainty [103].



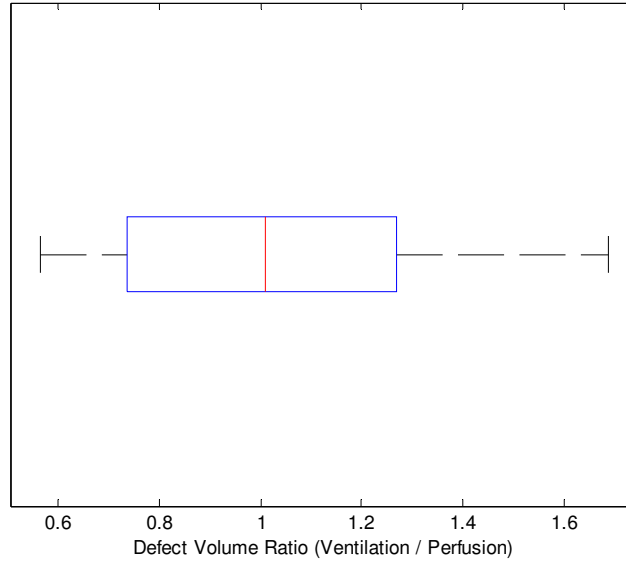
**Figure 4.5. Ventilation/perfusion spatial overlap indices.** Average ( $\pm$  standard deviation) DSC, NDSC<sub>1mm</sub>, and NDSC<sub>2mm</sub> between ventilation and perfusion hypo-functioning regions are shown for all cases included in this study.

Figure 4.6 shows a scatter plot of the average spatial overlap indices versus the corresponding volume of the reference perfusion defect region, normalized to the segmented volume of the affected lung. The square of Pearson's linear correlation coefficient was  $R^2 = 0.37$  ( $p = 0.060$ ),  $R^2 = 0.41$  ( $p = 0.048$ ), and  $R^2 = 0.50$  ( $p = 0.02$ ) for DSC, NDSC<sub>1mm</sub>, and NDSC<sub>2mm</sub>, respectively. The linear correlation coefficients and corresponding  $p$ -values suggest weak correlation of spatial overlap index with size of the corresponding perfusion defect region.



**Figure 4.6. Spatial overlap versus defect volume.** Average DSC, NDSC<sub>1mm</sub>, and NDSC<sub>2mm</sub> values for each case are shown versus the corresponding size of the reference perfusion defect region, normalized to the volume of the affected lung. Linear regression suggests weak correlation, bordering on statistical significance ( $p \approx 0.05$ ).

A boxplot is shown in figure 4.7 summarizing the ratio of ventilation/perfusion defect volumes for the set of patient cases. The figure demonstrates a trend towards V/Q volume ratios  $> 1.0$ , with average ratio equal to 1.05. One could postulate the reason for the relatively larger ventilation defect volumes could be attributable to the fundamental difference in acquisition settings between ventilation and perfusion images. The ventilation derived from 4D CT represents a snap-shot of the spatial distribution of ventilation corresponding to a single respiratory cycle, whereas the functional map obtained from  $^{99m}\text{Tc}$  MAA SPECT is acquired over a significantly longer time scale, on the order of 20 minutes, allowing gradual accumulation of the radiopharmaceutical within defect regions affected by partial obstruction by the gross disease. The time-averaged signal and relatively low spatial resolution from perfusion SPECT effectively contribute to blurring of functional boundaries, thereby reducing the overall defect volume.



**Figure 4.7. Mean ratios for V/Q defect volumes.** The ratio of mean ventilation defect volume to the corresponding reference SPECT perfusion volume is summarized in boxplot format for all cases. Average and median volume ratios were 1.05 and 1.01, respectively, suggesting a trend toward relatively larger pulmonary defect regions as depicted using ventilation from 4D CT.

#### 4.4 Discussion

In this study, we evaluated the spatial correspondence of hypo-functioning pulmonary regions determined from nuclear medicine  $^{99m}\text{Tc}$  MAA SPECT perfusion imaging and specific ventilation from 4D CT in a population of 10 lung cancer patients with radiographically demonstrated malignant airway stenosis. A previously reported [115] DIR algorithm referred to as Moving Least Squares Guided Local Optimization (MILO), developed specifically to overcome the computational challenges associated with DIR of breath-hold CT image pairs obtained as part of the NHLBI funded COPDgene study, was used in this work to provide the correspondence of lung tissue elements between component phases necessary to quantify specific ventilation from 4D CT. Spatial accuracy performance of the MILO algorithm was assessed using previously characterized reference images, with corresponding samples of manually identified landmark feature pairs, and found to be statistically indistinguishable from estimates of observer variance obtained by repeat registration (table 4.4). For self-consistency analysis of the calculated ventilation maps, tidal volumes obtained directly from 4D CT ventilation were compared to measurements obtained from image segmentation of the lung parenchyma voxels for each case; linear regression demonstrated excellent overall agreement (figure 4.3). To assess spatial correspondence between hypo-functioning pulmonary defect regions, raw ventilation and

perfusion images were first converted into percentile distribution images in which each voxel value was replaced according to the value given by the corresponding cumulative distribution function, scaled to the interval [0, 100]. Using the Dice similarity coefficient as the spatial overlap metric, mean DSC was determined for each case from the set of manually contoured ventilation defect regions provided by three independent observers. The raw DSC values were further normalized to account for variability in magnitude volume of the reference perfusion defect ROIs, and to provide numerical context for the measured overlaps in terms of fixed spatial uncertainties of 1.0 and 2.0 mm, or approximately 1-2 pixels in the transverse plane. Quantitative results indicate that on average, spatial overlap agreement between ventilation and perfusion defect regions was comparable to the threshold for agreement within 1-2 mm uncertainty.

In Chapter 3 of this work, we evaluated the spatial correlation between 4D CT ventilation and the clinical reference  $^{99m}\text{Tc}$  DTPA aerosol SPECT ventilation using treatment planning datasets for seven lung and esophagus cancer patients receiving thoracic radiotherapy. In that study, correlation was assessed among non-overlapping percentile mask regions representing fixed intervals of percentile distribution of ventilation. Percentile regions were selected in 20% intervals, resulting in binary masks for each case representing the spatial distribution of (1-20), (41-60), (61-80), and (81-100) percentile regions within the segmented lung volume. DSC values in that study were comparatively low ( $< 0.7$ ), owing to the presence of aerosol deposition artifacts within the segmented lung regions, which confounded the overall comparative evaluation. However, statistical analysis revealed significantly higher correlation between 4D CT and SPECT ventilation within the lowest functioning percentile mask regions; this finding was particularly evident for three cases with macroscopic defect regions due to malignant airway obstruction (figure 3.9). The  $^{99m}\text{Tc}$  MAA radiopharmaceutical utilized in this study for clinical SPECT perfusion imaging is not susceptible to airway deposition artifacts, suggesting the percentile images more accurately reflected the physiologic distribution of pulmonary function (perfusion) than in the previous work. The results of the present work support our previous findings, suggesting that 4D CT ventilation can be used to identify the spatial distribution of hypo-functioning pulmonary regions in malignant airway stenosis.

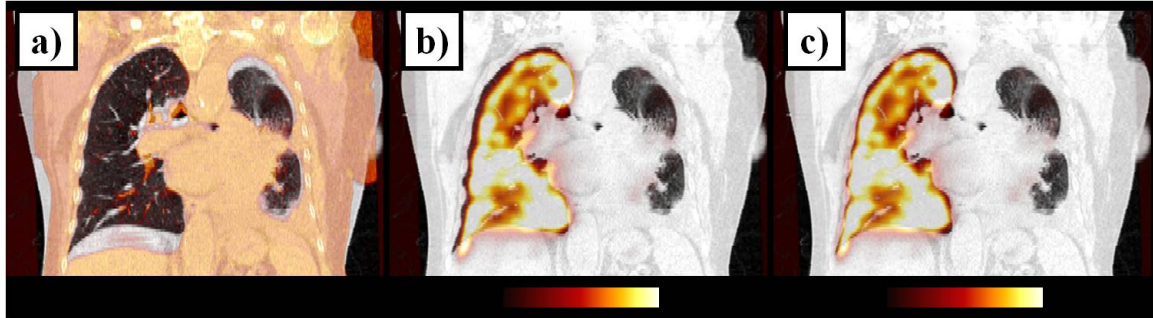
To date, these studies represent the most comprehensive physiologic validation of ventilation imaging from 4D CT using clinical reference lung function imaging in lung cancer patients. In 2007, Guerrero et al. demonstrated the use of ventilation obtained using a high resolution small animal CT scanner for quantifying the loss of pulmonary compliance in irradiated mice, 200 days following a single dose of 16 Gy  $\text{Co}^{60}$  hemi-thoracic irradiation [100]. In that study, irradiated mice had 60% lower mass-specific compliance than did the group of

controls, for compliance images obtained from 2 cm H<sub>2</sub>O and 10 cm H<sub>2</sub>O constant pressure CT images. In 2008, Reinhardt et al. compared measurements of Jacobian-based specific volume change to xenon CT measures of specific ventilation obtained from constant pressure images in five lung-healthy sheep [7]. In that study, the best match between average Jacobian and xenon CT ventilation was obtained for images acquired at 10 cm H<sub>2</sub>O and 15 cm H<sub>2</sub>O airway pressures, with average linear correlation coefficient  $R^2 = 0.73$ . Comparative evaluation was limited to approximately 3 cm axial coverage for each case. In a 2010 conference proceeding [120], Yamamoto et al. presented a single case study in which 4D CT ventilation was compared to SPECT ventilation and perfusion images, with low overall correlation demonstrated versus both ventilation (Spearman correlation coefficient = 0.03) and perfusion (Spearman correlation coefficient = 0.357).

NDSC<sub>1mm</sub> and NDSC<sub>2mm</sub> values reported in this study point to high spatial overlap agreement between 4D CT ventilation and SPECT pulmonary perfusion defect regions. Factors contributing to overall reduction in spatial correlation between methods include multimodal image registration, difference in acquisition settings, and motion and/or reconstruction artifacts present in one or both sets of images. For each case, the CT-to-CT fusion tool in the Pinnacle treatment planning system was used to align the T50 phase image from 4D CT with the CT component of the perfusion SPECT-CT study. The corresponding affine coefficient matrix was subsequently used to align and reconstruct the SPECT perfusion image to the spatial domain of the 4D CT for comparative evaluation with 4D CT ventilation, which is inherently co-registered to the T50 component phase image. The multimodal registration is subject to spatial alignment error due to potential differences in patient pose between the two images, as well as shifts in the nuclear medicine image that may have occurred between corresponding CT and SPECT perfusion acquisitions. To reduce the potential impact of these effects, visual inspection of the aligned nuclear medicine images was performed following affine registration, and manual translations imposed as necessary. An example of this is shown in figure 4.8.

Inherent differences in acquisition settings between nuclear medicine SPECT perfusion and 4D CT further contribute to overall reduction in spatial correlation between methods. In general, the SPECT image acquisitions occurred over a time scale on the order of 20 minutes to allow adequate accumulation of the radiopharmaceutical activity for functional image analysis. This is in contrast to the ventilation obtained from 4D CT, which represents the distribution of ventilation resulting from a single respiratory breathing phase. Thus, the imaged physiologic states are not entirely equivalent. The significantly longer acquisition of the SPECT perfusion images allows gradual accumulation of the radiopharmaceutical within defect regions distal to

partial obstruction that does not allow significant ventilation through one breathing cycle. The average ratio ( $> 1.0$ ) of ventilation / perfusion defect volumes found in this study supports this assertion (figure 4.7).

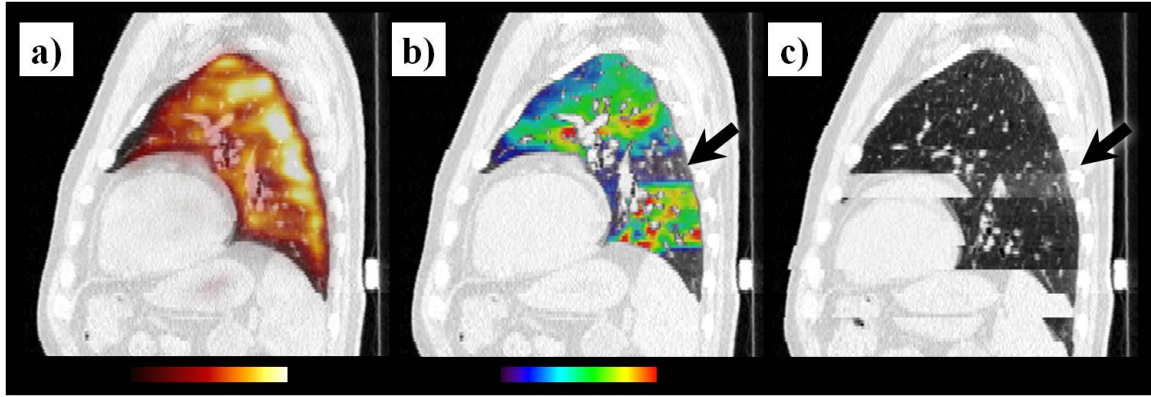


**Figure 4.8. Multimodal image alignment.** **a)** The CT component of the SPECT-CT perfusion study is shown in yellow overlain the corresponding T50 phase image from 4D CT, following affine registration as described above. **b)** The SPECT perfusion is similarly shown. Spatial alignment error is clearly visible along the right lung boundary. Note that no such misalignment is present in (a). **c)** SPECT perfusion is shown overlain the T50 component image, following manual adjustment of the volumetric image approximately 5 mm laterally to account for shifts in patient position between SPECT and CT acquisitions.

Finally, specific ventilation obtained from 4D CT is subject to degradation resulting from irregular breathing and subsequent acquisition artifacts. These artifacts (figure 4.9) cause the reconstructed CT values to violate our initial assumption that each component image volume represents a snap-shot of the respiratory cycle. Depending on the particular nature of the artifact, quantitative values for specific ventilation can be shifted artificially high or low, resulting in hot or cold spots that are not representative of the underlying pulmonary function. An example of this is shown in figure 4.9, which depicts a case in which a misplaced cine segment in the T00 phase image has resulted in a cold-spot banding artifact in the corresponding ventilation. Previously we showed that specific ventilation images derived from density change- or Jacobian-based approaches are subject to degradation from such artifacts [112]. In the present study, both T50 and T00 component phase images were made available to observers contouring on the percentile ventilation maps, in order that informed decisions could be made regarding the distinction between hypo-ventilated pulmonary regions versus artifact induced cold-spots arising from 4D CT artifacts. 4D CT acquisition methods to prevent such artifacts are needed. These findings point to the need for quality control practices that ensure informed interpretation of the functional maps prior to implementation into a specific clinical application, as well as improved 4D CT



acquisition and reconstruction techniques that reduce the overall frequency of 4D CT artifacts, which have been estimated previously to affect 90% of image acquisitions [121].



**Figure 4.9. Misplaced cine segment causes artifact on 4D CT ventilation.** **a)** SPECT pulmonary perfusion is shown overlain the T50 component phase image from 4D CT. **b)** The corresponding specific ventilation from 4D CT is also shown, with dark arrow showing cold-spot banding artifact resulting from **c)** misplaced cine segment in the T00 component image.

In previous work, Yaremko et al. evaluated in a treatment planning study with 21 NSCLC cases 4D CT ventilation images as an alternative means of radiotherapy treatment planning that might reduce complications from treatment [16]. In this approach, the ventilation images are used to identify regions of high function for conformal avoidance with IMRT. The ventilation avoidance plans were able to significantly reduce the dose to the highest functioning regions without increase in the calculated toxicity. More recent treatment planning studies have investigated the use of hyperpolarized  $^3\text{He}$  MRI ventilation images to provide functional avoidance [95, 96, 107]. Another 4D CT ventilation image guided conformal avoidance treatment planning study was reported by Yamamoto et al. [114]. Munawar et al. [97] reported a SPECT ventilation image guided conformal avoidance treatment planning study. Yin et al. [122] evaluated the impact of scatter and attenuation correction on SPECT perfusion images for radiotherapy planning. Functional volume segmentation was sensitive to those corrections. In each of these planning studies the approach is to identify the highest functioning pulmonary regions. In future work, we propose to identify regions which are hypo-functioning due to malignant airway stenosis as well, such that both high and low function regions will contribute to the inverse planning. To date, no clinical trial has tested the functional image guidance concept.



#### **4.5 Conclusion**

In this study, we evaluated the spatial correspondence of hypo-functioning pulmonary regions determined from nuclear medicine  $^{99m}\text{Tc}$  MAA SPECT perfusion imaging and specific ventilation from 4D CT in a population of 10 lung cancer patients with radiographically demonstrated malignant airway stenosis. Using the Dice similarity coefficient as the spatial overlap metric, we found high correlation between methods, comparable to the threshold for agreement within 1-2 mm uncertainty. These findings suggest the use of 4D CT ventilation for conformal avoidance as well as dysfunctional allowance in inverse planning for thoracic radiotherapy. This study is the first to quantitatively assess three-dimensional spatial correlation between clinically acquired SPECT perfusion and specific ventilation from 4D CT.

## Chapter 5

### Identification of the Independent Contribution of Pulmonary Perfusion to the Observed CT Image Intensities

We have shown in this work that the local fractional volume change due to inspiration can be quantified for CT images obtained at different respiratory states (i.e., “inhale” and “exhale”), according to the following equation:

$$\frac{\Delta V}{V_{air}^{exhale}} = \frac{F_{air}^{inhale} - F_{air}^{exhale}}{F_{air}^{exhale} (1 - F_{air}^{inhale})}, \quad (36)$$

where  $F_{air}$  is the fractional air content within the corresponding inhale and exhale sub-volumes. This is a familiar expression, from which an equivalent expression for the specific ventilation can be derived in terms of registered HUs between inhalation and exhalation breathing states (see Appendix). This formulation, which is based on the simple assumption that any pulmonary sub-volume can be modeled as a linear combination of water and air, implicitly assumes that any change in density between registered CT volume elements is the result of air volume change due to inspiration. In practice, when comparing uncorrected (see Chapter 3) tidal volume calculations to measurements obtained from image segmentation we find systematic underestimation of the calculated global volume change. This suggests that the difference in registered CT numbers, i.e.,  $\Delta HU = HU^{inhale} - HU^{exhale}$ , is smaller than would be expected based solely on consideration of the pulmonary ventilation. Previously it has been shown that the lung mass (as quantified from 4D CT) is cyclic with the respiratory phase [15]. This cyclic variation represents the changing perfusion in the pulmonary parenchyma that results from local mechanical factors, such as distension of the blood vessels, and from variation in respiratory-induced cardiac output [123, 124].

It can be shown that the regional lung density within a CT volume element can be expressed as:

$$\rho_{voxel}^{lung} = 1 + \frac{HU_{voxel}^{lung}}{1000}, \quad (37)$$

where the density  $\rho$  is in units of [g/ml]. Rearranging, we obtain:

$$HU_{\text{voxel}}^{\text{lung}} = 1000(\rho_{\text{voxel}}^{\text{lung}} - 1), \text{ for } [0 \leq \rho_{\text{voxel}}^{\text{lung}} \leq 1]. \quad (38)$$

Thus, the introduction of additional blood ( $\rho^{\text{blood}} \approx 1 \text{ g/ml}$ ) into a given pulmonary sub-volume has the effect of increasing the observed CT image intensity. This effect is not explicitly taken into account in our ventilation model, and is therefore a potential source of error, and the most likely reason for the systematically low calculations of the regional volume change. Therefore, it is of interest to isolate the contribution of blood to the observed change in HU, both to remove its effect on quantification of the regional volume change, as well as to potentially extract an independent measure of the pulmonary perfusion.

In Chapter 3 we showed that the local absolute volume change can be determined from 4D CT, independent of the observed image intensities:

$$\Delta V^{\text{Jac}}(\vec{x}) = 1 - J(\vec{x}), \quad (39)$$

where  $J(\vec{x})$  is the determinant of the Jacobian of the vector-valued spatial transformation function relating the underlying tissue elements between inhalation and exhalation breathing states. Substitution into equation (36) gives the following:

$$\frac{\Delta V^{\text{Jac}}}{V_{\text{air}}^{\text{exhale}}} = \frac{F_{\text{air}}^{\text{inhale}} - F_{\text{air}}^{\text{exhale}}}{F_{\text{air}}^{\text{exhale}} (1 - F_{\text{air}}^{\text{inhale}})}. \quad (40)$$

Given the Jacobian measure of absolute volume change, one can rearrange the above equation to infer the fraction of air within the inhalation state sub-volume:

$$F_{\text{air}}^{\text{inhale}} = \frac{F_{\text{air}}^{\text{exhale}} (V_{\text{air}}^{\text{exhale}} + \Delta V^{\text{Jac}})}{V_{\text{air}}^{\text{exhale}} + (F_{\text{air}}^{\text{exhale}} \Delta V^{\text{Jac}})}. \quad (41)$$

Consider the case where the sub-volume of interest is a discretized image voxel. Then the volume  $V_{\text{air}}^{\text{exhale}}$  is given simply as the product of the fraction of air at exhale and the voxel volume:

$$V_{\text{air}}^{\text{exhale}} = F_{\text{air}}^{\text{exhale}} V_{\text{voxel}}. \quad (42)$$

Substituting into equation (41) and simplifying, we obtain:

$$F_{air}^{inhale} = \frac{F_{air}^{exhale} V_{voxel} + \Delta V^{Jac}}{V_{voxel} + \Delta V^{Jac}}. \quad (43)$$

Recall that the initial fraction of air within the exhalation state voxel is given by:

$$F_{air}^{exhale} = -\frac{HU^{exhale}}{1000}. \quad (44)$$

Substitution into equation (43) yields an expression for the fractional air content within the inhalation state voxel:

$$F_{air}^{inhale} = \frac{\left(-HU^{exhale}/1000\right)V_{voxel} + \Delta V^{Jac}}{V_{voxel} + \Delta V^{Jac}}. \quad (45)$$

Thus, assuming deformable image registration between CT images corresponding to consecutive exhalation and inhalation breathing states (e.g., component phases from 4D CT), the Jacobian of the DIR spatial transformation function allows explicit quantification of the absolute regional volume change, independent of the underlying image intensity values. Based on the calculated  $\Delta V^{Jac}$  at any arbitrary lung position in the exhalation state image, one can determine the corresponding fraction of air in the inhale sub-volume, by virtue of equation (45).

Alternatively, we have shown that the fractional air content  $F_{air}^{inhale}$  can be determined directly from the inhale CT image intensities:

$$F_{air}^{inhale} = -\frac{\bar{H}_{VOI}^{inhale}}{1000}, \quad (46)$$

where  $\bar{H}_{VOI}^{inhale}$  is the average HU of all inhale voxels that map into a single exhale lung voxel, as determined by the DIR.

Equations (45) and (46) represent independent formulations of a single quantity, namely, the fraction of air in a lung volume of interest in the inhalation state that is related to its exhale counterpart by the DIR. Equation (45) is independent of the inhale CT image intensities,

depending primarily on the calculated spatial transformation. In contrast, equation (46) requires explicit knowledge of the inhale CT numbers. Both formulations require deformable image registration between the CT image pair.

In practice, for a given DIR solution, equations (45) and (46) will not give the same numerical answer, even in the ideal case where the DIR is exact. This is because the inhale HU are expected to encode the subtle density change that results from the difference in pulmonary blood perfusion between the inhalation and exhalation image states. Thus, in principle, the numerical discrepancy between equations (45) and (46) could itself be considered a direct measure of the impact of additional blood content on the observed inhale CT image intensity.

Consider the effect of blood perfusion as an unknown additive contribution to the observed Hounsfield Unit:

$$\bar{H}_{VOI}^{observed} = \bar{H}_{VOI}^{air} + \Delta\bar{H}_{VOI}^{blood}, \quad (47)$$

where  $\Delta\bar{H}_{VOI}^{blood}$  is the increase in HU that results from the change in blood content between inhale and exhale breathing states, and  $\bar{H}_{VOI}^{air}$  is the inhale HU that would be expected if the change in density at that position was due solely to ventilation. To derive an expression for  $\Delta\bar{H}_{VOI}^{blood}$ , first rewrite equation (46) in terms of the two unknown HU components:

$$\begin{aligned} F_{inhale,air}^{observed} &= \left(-\frac{1}{1000}\right) \bar{H}_{VOI}^{observed}; \\ F_{inhale,air}^{observed} &= \left(-\frac{1}{1000}\right) (\bar{H}_{VOI}^{air} + \Delta\bar{H}_{VOI}^{blood}); \\ F_{inhale,air}^{observed} &= \left(-\frac{\bar{H}_{VOI}^{air}}{1000}\right) + \left(-\frac{\Delta\bar{H}_{VOI}^{blood}}{1000}\right); \\ F_{inhale,air}^{observed} &= F_{air}^{inhale} - \left(\frac{\Delta\bar{H}_{VOI}^{blood}}{1000}\right). \end{aligned} \quad (48)$$

Define the quantity  $\lambda$  as the discrepancy between Jacobian- (equation (45)) and observed density-based (equation (48)) measures of the inhale fractional air content:

$$\lambda \equiv F_{inhale,air}^{Jacobian} - F_{inhale,air}^{observed} = F_{inhale,air}^{Jacobian} - \left( F_{air}^{inhale} - \left[ \frac{\Delta \bar{H}_{VOI}^{blood}}{1000} \right] \right);$$

$$\lambda = \left( F_{inhale,air}^{Jacobian} - F_{air}^{inhale} \right) + \left( \frac{\Delta \bar{H}_{VOI}^{blood}}{1000} \right).$$
(49)

Thus, in the ideal scenario (i.e.,  $F_{inhale,air}^{Jacobian} = F_{air}^{inhale}$ ), the contribution of image intensity related to change in blood content between inhalation and exhalation breathing states can be expressed simply in terms of the observed discrepancy  $\lambda$  in inhale fractional air content:

$$\Delta \bar{H}_{VOI}^{blood} = 1000 * \lambda .$$
(50)

Equation (50) can be interpreted as the change in average Hounsfield Unit within a volume of interest that is attributed to change in blood content between spatially registered inhalation and exhalation breathing states, given the quantifiable discrepancy  $\lambda$ , which is simply the element-wise numerical difference between equations (45) and (46).

## Chapter 6

### Summary

#### 6.1 Evaluation of Deformable Image Registration

A novel MATLAB-based interface application was developed with the specific intent to facilitate manual identification and subsequent registration of anatomical landmark features between multiple sets of volumetric medical image data (e.g., treatment planning 4D CT). Computer assistance tools provide for rapid feature localization, while still requiring that the human observer ultimately designate all feature correspondences. We have shown in this work that the APRIL software package can be used to generate large samples of expert-identified feature correspondences, with sufficiently low inter-observer variance in feature registration to facilitate quantitative spatial accuracy assessment of deformable image registration with a narrow range of uncertainty. Furthermore, we have provided guidelines based on statistical considerations for proper use of reference samples in DIR validation and comparative evaluation studies, to reduce the likelihood of erroneous interpretation of spatial accuracy measurements based on reference samples that are either too few in sample size, or insufficiently distributed throughout the anatomic target of interest to provide unbiased assessment of algorithm performance. To date, the methods developed in this study have led to improvement and application-specific optimization of three in-house developed DIR algorithms, which are continually refined to meet the high spatial accuracy requirement of our functional imaging application [98, 115, 125].

#### 6.2 Correlation of 4D CT Ventilation with $^{99m}\text{Tc}$ DTPA Aerosol SPECT Ventilation

Two calculation methods to produce ventilation images from 4D CT acquired without added contrast have been reported in the literature. The first, a density-based CT value method, quantifies three-dimensional functional images of the local specific ventilation using deformable image registration and the underlying CT density information. A second method performs the ventilation image calculation from the DIR result alone, using the Jacobian determinant of the deformation field to estimate the local volume changes resulting from ventilation. For each of these two approaches, variations on their implementation have been reported. In this study, two implementations of the Jacobian-based methodology were evaluated, as well as a single density change-based model, which we derive in the Appendix, for calculating the physiologic specific ventilation from 4D CT. In clinical practice,  $^{99m}\text{Tc}$ -labeled aerosol single photon emission computed tomography is the standard method used to obtain ventilation images in patients. In the present work, the distributions of ventilation obtained from the CT-based ventilation image calculation methods were compared with those obtained from the clinical standard SPECT

ventilation imaging. Seven patients with 4D CT imaging and standard  $^{99m}\text{Tc}$ -labeled aerosol SPECT-CT ventilation imaging obtained on the same day as part of a prospective validation study were selected. All 4DCT-based images of specific ventilation showed good linear correlation with image segmentation-based measures of the global resting tidal volume. Spatial correlation of the distribution of ventilation with the clinical SPECT was assessed using the Dice similarity coefficient, which showed statistically higher ( $p\text{-value} < 10^{-4}$ ) correlation between density-change based specific ventilation and the clinical reference than did either Jacobian-based implementation.

### **6.3 Correlation of Hypo-Functioning Pulmonary Regions in Malignant Airway Disease**

Abratt et al. showed in a 1990 retrospective study that delivery of radiotherapy through hypo-perfused pulmonary regions for lung cancer treatment was found to result in less pulmonary injury [1], concluding that less lung damage could be expected if there is decreased perfusion in the pulmonary regions irradiated. In this study, we hypothesized that hypo-ventilated regions determined from 4D CT will correlate with hypo-perfused pulmonary regions in lung cancer patients who have obstructing lesions. To test this hypothesis, treatment planning 4DCT images were obtained retrospectively for 10 lung cancer patients with radiographically demonstrated airway obstruction due to gross tumor volume. Each patient had also received a SPECT perfusion study within one week of the planning 4DCT, and prior to the initiation of treatment. Deformable image registration was used to map corresponding lung tissue elements between the extreme component 4DCT phase images, from which quantitative three-dimensional images representing the local pulmonary specific ventilation were constructed. Semi-automated segmentation of the percentile perfusion distribution was performed to identify regional defects distal to the known obstructing lesion, and similarly performed by multiple observers to delineate corresponding defect regions depicted on 4DCT ventilation. The normalized Dice similarity coefficient index was determined for each observer between SPECT and ventilation defect regions over the set of patient cases to assess spatial correlation.

Consistent with our prior results, all 4D CT-based measures of specific ventilation showed good linear correlation with image segmentation-based measures of the global resting tidal volume. Moreover, spatial correlation with SPECT pulmonary perfusion imaging was high, with average DSC, NSDC<sub>1mm</sub>, and NDSC<sub>2mm</sub> values 0.78 [range: (0.60-0.99)], 0.88 [range: (0.71-1.07)], and 0.99 [range: (0.79-1.16)], respectively. These findings confirm our hypothesis that hypo-ventilated regions determined from 4D CT will correlate with hypo-perfused pulmonary regions in lung cancer patients who have obstructing lesions.



#### 6.4 Evaluation of the Hypothesis

The hypothesis of this study was that quantitative images depicting regional specific ventilation from 4D CT would provide an accurate assessment of pulmonary function for use in image guided thoracic radiation therapy.

To evaluate the hypothesis, it was first necessary to develop a consistent framework for quantitative and objective spatial accuracy performance assessment of deformable image registration, which itself is central to the functional image calculations. This was accomplished in Specific Aim 1, in which both the practical feasibility and inherent statistical necessity of the use of large samples of corresponding feature pairs between images for this purpose was demonstrated for a cohort of clinically acquired treatment planning 4D CTs. Specific Aim 1 of this work presents a framework and corresponding novel software infrastructure for quantitative evaluation of DIR spatial accuracy. The results demonstrate that large landmark point sets provide an effective means for objective evaluation of deformable image registration with a narrow uncertainty range, and further suggests a practical strategy for quality assurance of DIR spatial accuracy in the routine clinical setting.

The evaluation tools established in the first aim allowed us to focus on the main problem of this work, the extraction of quantitative lung function information from 4D CT and their associated deformations. Presently, nuclear medicine SPECT is the standard method for obtaining clinical pulmonary ventilation images. Thus, in Specific Aim 2 we investigated the correlation between the spatial distributions of ventilation derived from 4D CT and  $^{99m}\text{Tc}$  DTPA aerosol SPECT in a population of lung and esophagus cancer patients, prior to the initiation of treatment. For the seven cases included in the study, spatial overlap analysis demonstrated the highest correlation among methods was in the lowest functioning percentile regions (i.e.,  $\leq 20\%$ ). This finding reveals that the greatest similarity in spatial distribution was within the hypo-ventilating regions; this was especially true for three cases with malignant airway disease causing airway obstruction (see figure 3.9). Furthermore, it was noted that more detailed quantitative comparative evaluation between ventilation imaging methods was confounded by the presence of artifacts on the SPECT ventilation images due to airway deposition of the  $^{99m}\text{Tc}$  DTPA aerosol. The deposition artifacts are a known drawback to  $^{99m}\text{Tc}$  DTPA aerosol SPECT imaging, and have been previously reported by others [2, 5].

Perfusion imaging, performed using  $^{99m}\text{Tc}$  MAA nuclear medicine SPECT, has previously demonstrated the heterogeneous distribution of lung function among lung cancer patients presenting for radiation therapy [45, 126, 127]. In one study of 50 lung cancer patients,

74% of patients were found to have perfusion defects adjacent to their primary lung malignancy [128]. However, the complexity of the procedure and the resultant poor image resolution has limited its use in radiotherapy treatment planning. In Specific Aim 3, we investigated the correlation of regional hypo-ventilation derived from 4D CT with hypo-perfused pulmonary regions on  $^{99m}\text{Tc}$  MAA nuclear medicine SPECT perfusion. Based on the findings in Specific Aim 2, the evaluation was limited to a sub-population of lung cancer patients with radiographically demonstrated airway obstruction due to their gross tumor volume. Spatial overlap analysis for the 10 cases included in the study yielded mean  $\text{NDSC}_{2\text{mm}}$  values between volumetric ventilation and perfusion defect regions with mean and range of 0.99 (0.79-1.16). These findings suggest high spatial correlation between pulmonary perfusion and ventilation macroscopic defect regions in lung cancer patients with malignant airway stenosis.

Thus, using high spatial accuracy deformable image registration to facilitate lung function quantification, we have demonstrated in this work high spatial correlation of 4D CT ventilation with the clinical reference nuclear medicine SPECT ventilation and perfusion imaging, within local defect regions attributable to malignant airway stenosis. These findings limit the scope of the original hypothesis, but nonetheless demonstrate it to be true in this well-defined sub-population of lung cancer patients receiving thoracic radiotherapy.

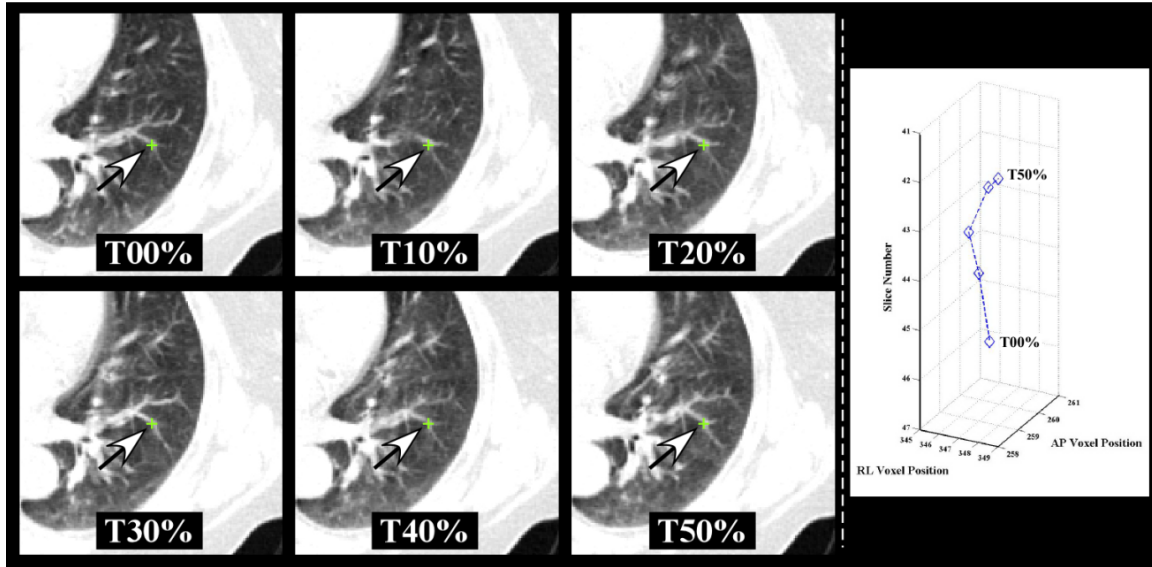
## **6.5 Ongoing and Future Applications**

### *6.5.1 The DIR-Lab Online Database*

Expert determined landmark correspondences have become a widely adopted reference for evaluating DIR accuracy for lung image data, though there has been great variability in their use. In this study we have presented a framework for objective evaluation of thoracic deformable image registration spatial accuracy, based on the use of large samples of expert-determined landmark feature pairs between volumetric images as a reference for spatial accuracy measurements (see Chapter 2). The DIR-Lab (<http://www.dir-lab.com>) is a website born out of this work, which was developed specifically to provide a comprehensive common dataset to investigators broadly associated with the field of deformable image registration who would like to evaluate their own algorithms, models, implementations, etc., using previously reported and statistically characterized reference datasets. Formally launched in March 2009, the DIR-Lab website makes publically available a repository of reference image data and image analysis software for quantitative clinical evaluation of deformable image registration.

Initially, five patient cases were retrospectively selected, and annotated with reference landmark positions identified on the maximum inspiration and expiration component phase

images from treatment planning 4D CT, as described in Chapter 2, and posted online for public download. Subsequent work by Castillo et al. [98] on the development of fully 4D spatio-temporal image registration algorithms compelled the need for reference information on intermediate component phase images, prompting additional manual propagation of subsets of feature points for each case across the expiration phase images (i.e., T50: T40: T30: T20: T10: T00). The process is illustrated in figure 6.1.



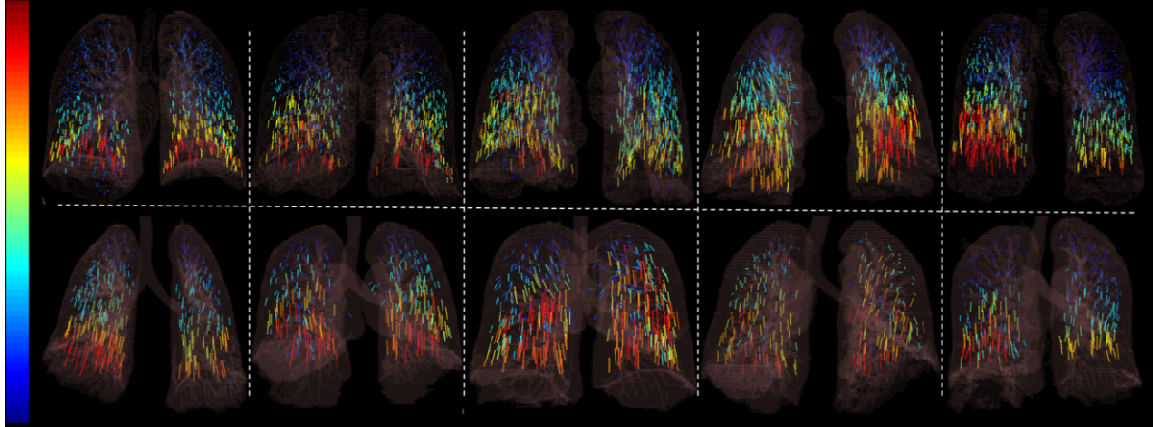
**Figure 6.1. 4D landmark trajectories.** Four dimensional landmark point sets were utilized to test the adequacy of polynomial trajectory models in [98]. **a)** The 4D CT image sets consisted of the six images spanning the expiratory phases from maximum inhalation (T00) to maximum exhalation (T50). Each landmark feature was identified (arrow) for phases T00 through T50 as shown. **b)** A sample 4D trajectory of the landmark point depicted in (a) is plotted. Note that the T30 and T40 positions overlay each other.

In addition to the original patient sets, five additional clinically acquired 4D CTs were selected from the M. D. Anderson patient database and similarly annotated with reference features on the expiration phase component images. The set of 10 reference 4D CTs was utilized in the study by Castillo et al. to evaluate the spatial accuracy performance of their novel implementation, and similarly added to the online database. The current list of reference image data available through the DIR-Lab website is shown below:

| Case Number | # Landmarks | Avg (SD) Displacement | Avg (SD) Observer Error | Landmarked Phases |
|-------------|-------------|-----------------------|-------------------------|-------------------|
| 1           | 1280        | 4.01 (2.91)           | 0.85 (1.24)             | [T00 : T50]       |
| 2           | 1487        | 4.65 (4.09)           | 0.70 (0.99)             | [T00 : T50]       |
| 3           | 1561        | 6.73 (4.21)           | 0.77 (1.01)             | [T00 : T50]       |
| 4           | 1166        | 9.42 (4.81)           | 1.13 (1.27)             | [T00 : T50]       |
| 5           | 1268        | 7.10 (5.14)           | 0.92 (1.16)             | [T00 : T50]       |
| 6           | 419         | 11.10 (6.98)          | 0.97 (1.38)             | [T00 : T50]       |
| 7           | 398         | 11.59 (7.87)          | 0.81 (1.32)             | [T00 : T50]       |
| 8           | 476         | 15.16 (9.11)          | 1.03 (2.19)             | [T00 : T50]       |
| 9           | 342         | 7.82 (3.99)           | 0.75 (1.09)             | [T00 : T50]       |
| 10          | 435         | 7.63 (6.54)           | 0.86 (1.45)             | [T00 : T50]       |

**Table 6.1. DIR-Lab reference database.** A summary of the available DIR-Lab reference case data is presented. All distances and error measurements are presented in units of millimeters. SD: standard deviation.

For each case, the “Observer Error” column shows the combined mean (and pooled standard deviation) repeat registration error for the set of three independent readers, as described in Chapter 2. The data show that the reference pulmonary features can be consistently and reliably identified, sufficient to evaluate DIR spatial accuracy performance with a narrow range of uncertainty. Figure 6.2 shows vector plots of the landmark displacement fields between maximum inhalation and exhalation phase images for the set of ten cases in anterior projection. As described previously, the sampled feature points are sufficiently distributed to illustrate substantially heterogeneous spatial distributions of tissue motion within each of the lung volumes. The variability in motion characteristics and disease states is crucial to any such online database, ensuring representation of as wide a range as possible of images encountered in clinical practice.



**Figure 6.2. DIR-Lab reference data renderings.** Manually determined displacement vectors are shown in anterior projection for the set of ten DIR-Lab reference 4D CTs. Characteristics of the landmark features and associated observer variances are provided in table 6.1.

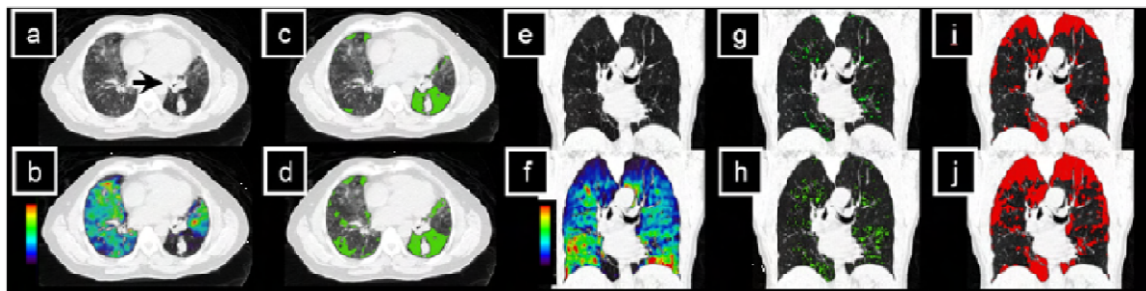
Public interest in the website has grown steadily since its initial inception, averaging 75 independent visits per day in the month of February 2011. This is up from just 8 visits per day maintained during the opening month of March 2009. Visitors to the website are encouraged to register freely online in order to access the reference datasets and associated software content. Currently, 80 research groups from around the world have registered, and regularly receive research and technical support. Registered users represent a cross-section of academic, commercial, and government interests, and come from various countries including the United States, Canada, France, Germany, Italy, Egypt, India, The Netherlands, Australia, and Japan. All registered users have the option of reporting spatial accuracy results back to the DIR-Lab for public posting on the web, along with reference to the published work in which the spatial accuracy results are formally reported. Presently, data are posted for four peer-reviewed publications [98, 99, 125, 129], and can be viewed at the dedicated results portion of the website: <http://www.dir-lab.com/Templates/Results.html>.

#### 6.5.2 CT Ventilation for Identifying Novel Imaging Biomarkers in COPD

Chronic Obstructive Pulmonary Disease (COPD), a collection of lung diseases associated with expiratory airflow obstruction, is the fourth leading cause of death in the United States. Clinically, COPD is diagnosed on the basis of airflow obstruction that is not fully reversible after the administration of inhaled bronchodilators. Two predominant disease phenotypes are apparent on biopsy emphysema and bronchiolitis (or small airway disease). The distribution of emphysema, the loss of lung parenchyma, can be readily identified with quantitative computed

tomography. Air-trapping due to bronchiolitis is a diagnostic challenge, often diagnosed only after biopsy. Current CT image analysis methods for COPD evaluation focus on the separate anatomic evaluation of exhale- and inhale-breath hold (BH) CT images [130]. In future work, we propose to evaluate the BH image pairs simultaneously using high spatial accuracy deformable image registration to link the two. Our goal is to identify regions of poor ventilation or air-trapping due to both emphysema and bronchiolitis. When combined with percentile point analysis, this method can identify the relative contributions of bronchiolitis and emphysema, each to lung function impairment in a single diagnostic setting.

The COPDgene study ([www.copdgene.org](http://www.copdgene.org)) is a National Heart Lung Blood Institute (NHLBI) funded cross-sectional study designed to discover what genetic factors contribute to the development of COPD. Over 5,000 patients have been recruited and 100 patients are added per week with a recruitment goal of 12,000 patients. Each patient included in the study received exhale BH and inhale BH-CT imaging, as well as pulmonary function testing. We will utilize the data from this study retrospectively to discover new imaging biomarkers of air-trapping due to bronchiolitis. Our biomarkers will utilize DIR to link the BH image pairs to obtain estimates of density change due to ventilation (see figure 6.3). Those regions with air trapping will have nearly constant CT values between the expiratory and inspiratory images and those areas with emphysema will have low CT values (e.g.,  $< -910$  HU). Imaging biomarkers will be devised using models based on the exhale HU, inhale HU, and ventilation values. Data-splitting will produce a set of data from which to develop the model and a separate test data set to evaluate the model. The model will identify bronchiolitic and emphysematous air trapping on a voxel by voxel basis. Correlation of the segmented regions and their properties with pulmonary function tests will be evaluated.



**Figure 6.3. Imaging biomarkers for COPD.** a) Demonstrates large airway obstruction due to NSCLC bronchial involvement causing air trapping. A lesion involving the segmental bronchial branch causes airway obstruction. The image shown was obtained from the maximum exhalation phase image from a 4D

CT set. **b)** The corresponding 4D CT-derived ventilation is shown overlain, and the lowest **c)** 15<sup>th</sup> and **d)** 30<sup>th</sup> percentile ventilation was used to identify air trapping regions (shown in green). **e)** An exhale BH CT image from a COPDgene study patient. **f)** The corresponding CT-derived ventilation is shown overlain. Emphysematous regions identified from **g)** -950 HU, and **h)** -910 HU binary masks (shown in green). The lowest **i)** 15<sup>th</sup> and **h)** 30<sup>th</sup> percentile of ventilation was used to identify air trapping (shown in red). Note the difference in distribution between the emphysematous and air-trapping regions.

Patient image sets associated with the COPDgene protocol present a unique set of technical challenges for spatially accurate deformable image registration, distinct from the characteristics of the treatment planning 4D CTs utilized in the present work. Challenges to spatially accurate registration of the maximum effort BH image pairs include large tissue displacements, large change in density and CT value, extreme difference in image noise due to the (1:4) ratio in acquisition tube current between exhale and inhale images, and the changes in anatomic shape of the vasculature due to the large volume change. To our knowledge, there are no published reports demonstrating a DIR formulation capable of achieving high spatial accuracy registration of the COPDgene data.

Recently, Castillo et al. [115] developed a novel algorithm, the Moving Least Squares Guided Local Optimization (MILO) method, designed specifically for this task, where preliminary accuracy assessments have been made using the methods described in this work. Five maximum effort BH-CT image pairs were selected from COPDgene study cases. Each patient had received CT imaging of the entire thorax in the supine position at normal expiration and maximum effort full inspiration. The CT imaging was performed with a GE VCT 64-slice scanner (GE Healthcare Technologies, Waukesha, WI) with a pitch of 1.375 mm, speed of 13.75 mm per rotation, 120 kV<sub>p</sub>, 0.5 sec per rotation, 400 mA per rotation for inhale BH, and 100 mA per rotation for exhale BH. The images used were reconstructed using a high resolution reconstruction algorithm (BONE), with the lung diameter setting the field of view, and with 2.5-mm slice spacing.

Measurements of DIR spatial accuracy were performed using manually identified sets of prominent anatomical landmark feature pairs identified across the maximum inhalation and exhalation BH-CT image pairs. Source feature points were selected systematically on the five test image pairs by an expert in thoracic imaging, beginning at the apex of the lung. Points were selected with an initial goal of >5 feature points for each lung per axial image slice. This approach ensured the collection of >600 validation point pairs for each case distributed throughout the lungs. Following feature selection for a given case, all landmark pairs were visually reviewed by the primary reader a second time and the locations adjusted on the exhale

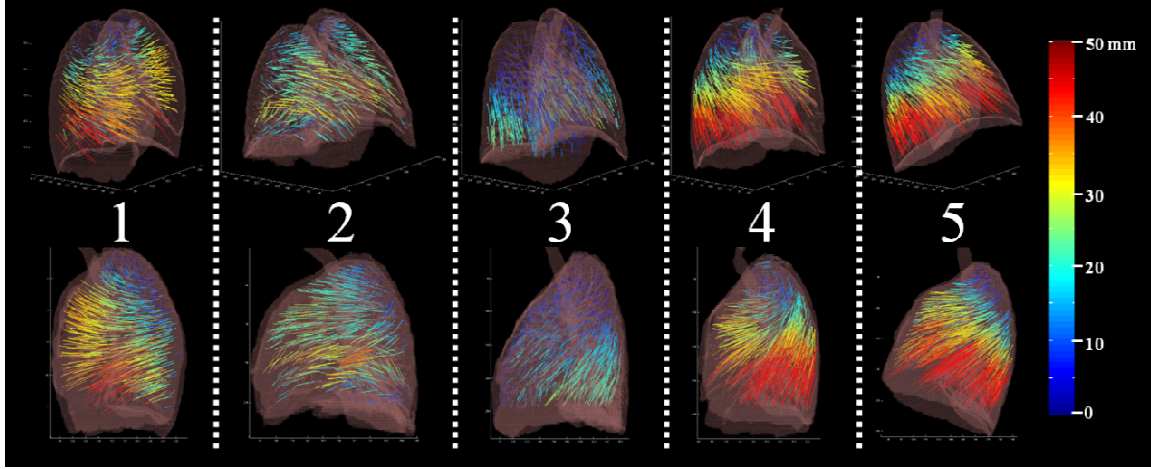
image if necessary. The verification step was a required part of the initial registration process performed by the primary reader. A subset of 150 points was re-registered by the primary reader, to estimate intra-observer variance, and by two secondary readers, to estimate inter-observer variance. For each of the 150 points given on the inhale BH-CT the re-registration process consisted of finding the corresponding point on the exhale BH-CT. Characteristics of the COPD reference datasets are given in table 6.2. Pending publication of the evaluation study in the peer-reviewed literature, public access to the reference dataset will be made available through the DIR-Lab online database.

| Case   | Image Dims  | Voxel Dims       | # Landmarks | Avg Displacement | Obs. Error  |
|--------|-------------|------------------|-------------|------------------|-------------|
| copd-1 | 512×512×121 | 0.625×0.625×2.50 | 773         | 25.90 (11.57)    | 0.65 (0.73) |
| copd-2 | 512×512×102 | 0.645×0.645×2.50 | 618         | 21.85 (6.58)     | 1.06 (1.51) |
| copd-3 | 512×512×126 | 0.652×0.652×2.50 | 1172        | 12.29 (6.39)     | 0.58 (0.87) |
| copd-4 | 512×512×126 | 0.590×0.590×2.50 | 786         | 30.90 (13.49)    | 0.71 (0.96) |
| copd-5 | 512×512×131 | 0.647×0.647×2.50 | 1029        | 30.90 (14.05)    | 0.65 (0.87) |

**Table 6.2. COPD image and reference data characteristics.** Image and voxel dimensions are shown for the set of five COPD reference datasets. Also shown is the number of reference landmarks for each case, along with the corresponding average (and standard deviation) landmark displacement. Estimates of observer variance in landmark registration were obtained by repeat registration and are also shown as mean (and standard deviation), combined for the set of multiple observers. All measurements of distance are reported in units of millimeters.

Figure 6.4 shows the set of reference displacement vector field projections for each case. Oblique and lateral aspects are shown at top and bottom, respectively. Transparent isosurfaces of the maximum effort inhale BH-CT are overlain the displacement field. The color scale shown at right is fixed for each projection, illustrating the variability in motion field characteristics among the five patient cases.





**Figure 6.4. COPD reference displacements.** Landmark point pairs were selected on exhale and inhale BH-CT images from five COPDgene cases. The number of pairs was 773, 618, 1172, 786, and 1029 for each case, respectively. A vector field projection is shown for each case using the color map for displacement length (0 to 50 mm), shown at right. Those vectors with  $> 50$  mm displacement are mapped to the same color as the 50 mm displacements. The range of displacements varies by patient case, reflecting the extent of their COPD disease.

Preliminary measurements of the MILO spatial accuracy obtained over the five COPDgene patient sets are summarized in table 6.3. Mean (and standard deviation) Euclidean magnitude errors ranged from 0.96 (1.10) to 2.47 (3.67) mm. For the combined set of COPDgene data ( $N = 1029$  reference measurements), mean spatial registration error was 1.27 (1.95) mm for MILO DIR.

| Case #                | Algorithm     | RL (mm)            | AP (mm)              | SI (mm)              | 3D (mm)              |
|-----------------------|---------------|--------------------|----------------------|----------------------|----------------------|
| <b>copd-1</b>         | <i>MILO</i>   | 0.40 (0.58)        | 0.80 (2.01)          | 0.62 (1.49)          | 1.38 (2.42)          |
|                       | <b>No DIR</b> | <b>4.31 (3.19)</b> | <b>23.66 (11.27)</b> | <b>7.39 (5.88)</b>   | <b>25.90 (11.57)</b> |
| <b>copd-2</b>         | <i>MILO</i>   | 0.77 (1.22)        | 1.24 (2.12)          | 1.42 (3.07)          | 2.47 (3.67)          |
|                       | <b>No DIR</b> | <b>3.35 (2.46)</b> | <b>19.68 (6.92)</b>  | <b>6.22 (5.44)</b>   | <b>21.85 (6.58)</b>  |
| <b>copd-3</b>         | <i>MILO</i>   | 0.38 (0.55)        | 0.44 (0.63)          | 0.40 (0.94)          | 0.98 (1.06)          |
|                       | <b>No DIR</b> | <b>2.11 (1.67)</b> | <b>6.11 (3.50)</b>   | <b>9.70 (6.41)</b>   | <b>12.29 (6.39)</b>  |
| <b>copd-4</b>         | <i>MILO</i>   | 0.42 (0.51)        | 0.50 (0.63)          | 0.41 (1.17)          | 1.05 (1.23)          |
|                       | <b>No DIR</b> | <b>5.75 (4.39)</b> | <b>17.67 (10.25)</b> | <b>21.20 (14.75)</b> | <b>30.90 (13.49)</b> |
| <b>copd-5</b>         | <i>MILO</i>   | 0.39 (0.53)        | 0.47 (0.76)          | 0.35 (0.88)          | 0.96 (1.10)          |
|                       | <b>No DIR</b> | <b>4.46 (3.44)</b> | <b>24.48 (11.31)</b> | <b>15.83 (11.94)</b> | <b>30.90 (14.05)</b> |
| <b>COPD-Composite</b> | <i>MILO</i>   | 0.45 (0.68)        | 0.63 (1.29)          | 0.57 (1.55)          | 1.27 (1.95)          |
|                       | <b>No DIR</b> | <b>3.88 (3.10)</b> | <b>17.52 (9.02)</b>  | <b>12.31 (9.69)</b>  | <b>23.76 (10.94)</b> |

**Table 6.3. MILO registration error summary.** Mean (and standard deviation) spatial registration errors are shown for the set of five COPDgene reference datasets. The unregistered landmark displacements (i.e., “No DIR”) are also shown for reference.

Demonstrating high spatial accuracy performance of the MILO DIR algorithm in the context of BH-CT images obtained from the COPDgene database is a necessary prerequisite for pursuing novel imaging biomarkers in COPD using the methods developed in this work. In future work, these preliminary assessments will be supplemented with measurements obtained using additional reference cases, which will also be added to the DIR-Lab online database. Bulk processing of the COPDgene data can then take place as described above, in order to assess the impact of the functional imaging technologies described in this work on diagnosis and phenotyping in COPD.

### 6.5.3 A Ventilation-based Strategy for DIR Quality Control and Assurance

There is a lack of consistency among evaluation strategies for DIR algorithm performance that confounds interpretive comparison of published results. In this work, we have devised a statistical framework for the comparison of DIR algorithms and demonstrated its validity. Currently, however, it is not clear that the selection of such large validation landmark sets could ultimately prove feasible for application to routine QA assessment of DIR. Prudent and responsible clinical implementation of deformable image registration and its associated applications will require a quality assurance check of each calculation prior to its use. Any such QA check should be fast and automated, to impart as little impact as possible on clinical workflow. This necessity for speed precludes the use of large landmark samples for real-time QA purposes.

For medical applications, the goal in applying DIR is to obtain an accurate spatial registration of the underlying anatomy. QA of DIR results should provide an estimate of the registration spatial accuracy for each use. Few studies addressing this problem exist in the literature. In previous work, Zhong et al.[41] proposed a finite-element-based metric for detecting displacement fields that violate continuum mechanics. Bender et al. [131] proposed the use of consistency metrics for QA assessment of DIR. The consistency of an algorithm can be determined, for example, by calculating deformation maps in both forward and reverse directions, and subsequently applying them to an image. The authors propose that if an algorithm is not inverse consistent, then the final image will not match the original unmodified image, indicative of regional mis-registration. To date, no validation studies have been performed demonstrating the correlation of proposed QA metrics with registration spatial accuracy or the presence of registration artifacts.

Based on findings reported in the present work, we propose a DIR quality assurance metric derived from two independent methods for calculating 4DCT ventilation. The fractional volume change is given by the absolute change in volume, normalized by the initial gas content within the exhalation state image voxels. We have shown that the fractional volume change can be determined explicitly from registered HU between CT images corresponding to the inhale/exhale breathing states:

$$\frac{\Delta V}{V_{exhale}^{air}} = 1000 \frac{\bar{H}_{inhale}^{VOI} - HU_{exhale}}{HU_{exhale} (1000 + \bar{H}_{inhale}^{VOI})}. \quad (51)$$

Alternatively, the magnitude volume change  $\Delta V$  can be determined directly from the determinant of the Jacobian matrix  $J(\vec{x})$  of the calculated DIR displacement field:

$$\Delta V(\vec{x}) = 1 - abs[J(\vec{x})]. \quad (52)$$

For a given pair of component phase images from 4D CT, equations (51) and (52) provide quantitative images depicting the local volume change due to inspiration, both of which inherently depend on the deformable registration linking the two images.

Consider the ideal scenario in which a perfect registration exists between the inhale/exhale breathing states. In that special case, the quantity  $\Delta V$  can be used to recover the exhale image data exactly, via the initial gas content  $V_{exhale}^{air}$  given in equation (51). Recall that the

gas volume can be expressed as the product of the voxel volume and fractional air content within that volume:

$$V_{exhale}^{air} \equiv V_{voxel} F_{exhale}^{air} = V_{voxel} \left( -\frac{HU_{exhale}}{1000} \right). \quad (53)$$

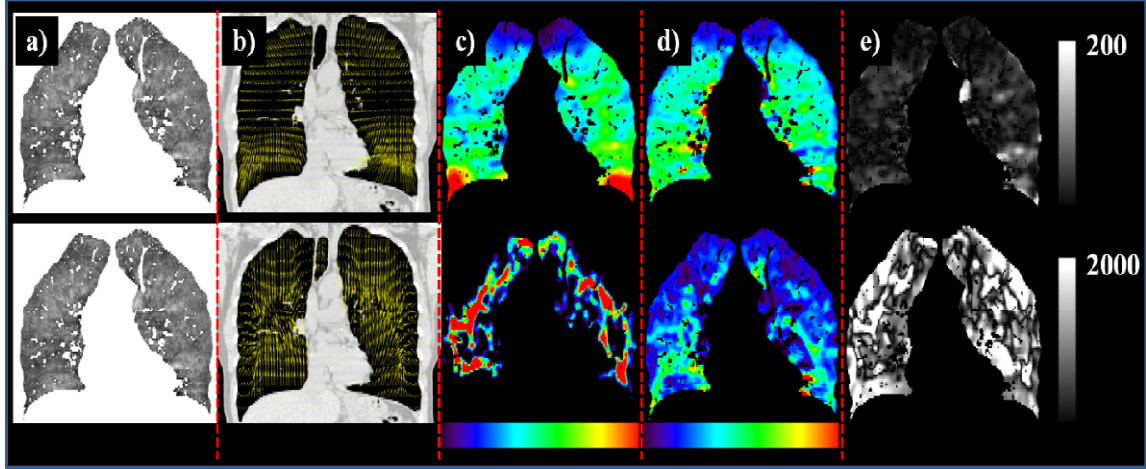
Division of equation (52) by equation (51) then gives the following:

$$\frac{\Delta V}{(\Delta V/V_{exhale}^{air})} = V_{voxel} \left( -\frac{HU_{exhale}}{1000} \right). \quad (54)$$

Solving for the exhale HU, we obtain:

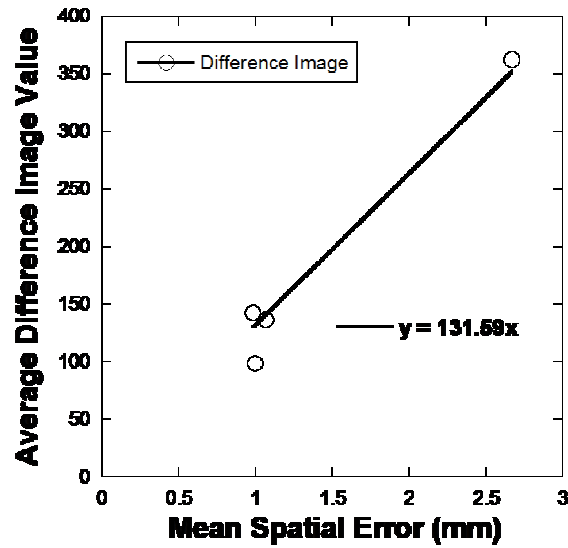
$$HU_{exhale} = -1000 \frac{Q}{V_{voxel}}, \quad (55)$$

where the quotient  $Q = \frac{\Delta V}{(\Delta V/V_{exhale}^{air})}$  is given by element-wise division of the quantitative images representing the magnitude, and fractional volume changes. Thus in the ideal case, equation (55) can be used to recover the exhale CT image data exactly. In practice, statistical noise in the quantitative CT numbers, as well as image acquisition and reconstruction artifacts contribute as potential sources of error. However, due to preprocessing of the input images, and spatial image smoothing that is inherent to the ventilation calculations, their contribution is likely negligible relative to spatial registration errors in the DIR (see Chapter 3). Thus, in principle, the quantitative estimate of the exhale CT given by equation (55) can be compared with the measured exhale CT image as a quality check of the DIR result. Figure 6.5 demonstrates the process for an example case. Two independent algorithms were used to register case *copd-1* given in table 6.2, yielding both high and low spatial accuracy outputs, with corresponding measured mean (and standard deviation) errors 1.38 (2.42) mm and 25.28 (11.08) mm, respectively.



**Figure 6.5. Automated QA/QC for DIR.** **a)** A coronal section through the pulmonary masked and smoothed exhale CT image. **b)** The corresponding inhale CT with displacement vectors overlain for high (top panel) and low (bottom panel) spatial accuracy DIR. **c)** Magnitude volume change determined from the Jacobian matrix. **d)** Specific ventilation from corresponding inhale/exhale HU. **e)** The difference image, the estimated exhale CT image subtracted from the original exhale image, where the estimate is made from equation(55). High intensity regions represent underlying spatial registration error, invalidating the equivalence of the Jacobian and HU ventilation calculations. Note the intensity scale differences of the high [0, 200] and low [0, 2000] spatial accuracy difference images.

Visually, the difference images are seen to record large magnitude intensities corresponding to poor spatial accuracy registration, with relatively low and homogeneous intensities corresponding to high spatial accuracy registration. Using the high spatial accuracy MILO algorithm, this process was repeated for four of the COPDgene cases given in table 6.2, and the average magnitude intensity difference plotted versus the corresponding mean spatial error. The result is shown in figure 6.6.



**Figure 6.6. Spatial error versus difference image QA metric.** The spatial error for the MILO DIR on four COPDgene cases with >600 landmark point pairs each. The ventilation estimated exhale image was constructed and subsequently subtracted from the measured exhale CT image, and the mean difference taken as the QA metric. The plot demonstrates correlation of the proposed QA metric with spatial error.

For the cases depicted, the figure demonstrates correlation of the proposed DIR quality assurance metric with the measured spatial error. In future work, we propose to create an image library of extreme phase (inhale/exhale) 4DCT image pairs, with manually registered (>150 per image pair) landmark points using the APRIL interface [99]. This test set library will consist of 50 representative lung cancer cases, each with 150 landmark pairs per case. The library, which will be made freely available through the DIR-Lab website, will form the standard for DIR spatial accuracy assessment. Using this library we will evaluate the correlation of the proposed DIR QA metric, as well as existing proposals from the literature, with measured spatial accuracy.

## Chapter 7

### Appendix: Determination of Specific Ventilation from Registered CT Hounsfield Units between Inhalation and Exhalation Breathing States

The ventilation methodology is based on the assumption that any given region of lung can be considered a combination of “air” and “tissue.” Let the volume  $V_1$  of a given region of interest (ROI) in the lower pressure, or exhale image be given as:

$$V_1 = V_{Air}^{Ex} + V_{Tissue} , \quad (A.1)$$

where  $V_{Air}^{Ex}$  is the volume of air in the region of interest and  $V_{Tissue}$  the corresponding volume of “tissue” in the same region. By definition, we can write an expression for the fraction of air contained within the given volume as:

$$F_{1,Air} = \frac{V_1 - V_{Tissue}}{V_1} . \quad (A.2)$$

Similarly, let the volume  $V_2$  of the corresponding region in the higher pressure, or inhale CT image be written:

$$V_2 = V_{Air}^{In} + V_{Tissue} , \quad (A.3)$$

with fractional air content given by:

$$F_{2,Air} = \frac{V_2 - V_{Tissue}}{V_2} . \quad (A.4)$$

The local fractional change in air content due to inspiration is the specific ventilation, and can be expressed:

$$sVent_{HU} \equiv \frac{\Delta V}{V_{Air}^{Ex}} = \frac{V_{Air}^{In} - V_{Air}^{Ex}}{V_{Air}^{Ex}} . \quad (A.5)$$

Substitution of equations (A.1) and (A.3) into the expression for specific ventilation and simplifying yields:

$$sVent_{HU} = \frac{V_2 - V_1}{V_1 - V_{Tissue}}. \quad (A.6)$$

Multiplying top and bottom of equation (A.6) by  $V_{Tissue}$  and distributing terms in the numerator yields:

$$sVent_{HU} = \frac{V_2 V_{Tissue} - V_1 V_{Tissue}}{V_{Tissue} (V_1 - V_{Tissue})}. \quad (A.7)$$

We then add the quantity 0 to the numerator in the form  $(V_1 V_2 - V_1 V_2)$ , and rearrange to arrive at the following:

$$sVent_{HU} = \frac{V_1 V_2 - V_1 V_{Tissue} - V_1 V_2 + V_2 V_{Tissue}}{V_{Tissue} (V_1 - V_{Tissue})}. \quad (A.8)$$

Simultaneously factoring terms in the numerator and multiplying denominator by unity in the form  $(V_1 V_2 / V_1 V_2)$  gives the following:

$$sVent_{HU} = \frac{V_1 (V_2 - V_{Tissue}) - V_2 (V_1 - V_{Tissue})}{V_1 V_2 \left( \frac{V_1 - V_{Tissue}}{V_1} \right) \left( \frac{V_{Tissue}}{V_2} \right)}. \quad (A.9)$$

Again multiplying by unity, this time in the form of the complex fraction  $\left( \frac{1/V_1 V_2}{1/V_1 V_2} \right)$  yields:

$$sVent_{HU} = \frac{\left( \frac{V_2 - V_{Tissue}}{V_2} \right) - \left( \frac{V_1 - V_{Tissue}}{V_1} \right)}{\left( \frac{V_1 - V_{Tissue}}{V_1} \right) \left[ \frac{V_{Tissue}}{V_2} \right]}. \quad (A.10)$$



Finally, we add the quantity 0 to the numerator of the bracketed term in equation (A.10) in the form  $(V_2 - V_2)$ , and simplify the resultant expression to obtain:

$$sVent_{HU} = \frac{\left( \frac{V_2 - V_{Tissue}}{V_2} \right) - \left( \frac{V_1 - V_{Tissue}}{V_1} \right)}{\left( \frac{V_1 - V_{Tissue}}{V_1} \right) \left( 1 - \frac{V_2 - V_{Tissue}}{V_2} \right)}. \quad (A.11)$$

Substitution of equations (A.2) and (A.4) into equation (A.11) then yields the simplified expression for  $sVent_{HU}$  in terms of the fractional air content within corresponding high and low pressure states of a given ROI:

$$sVent_{HU} = \frac{F_{2,Air} - F_{1,Air}}{F_{1,Air} (1 - F_{2,Air})}. \quad (A.12)$$

Thus, to arrive at equation(16)we need only substitute an expression for the fractional air content within an arbitrary lung ROI in terms of the CT Hounsfield Unit (HU). Note that in general,  $size(ROI_2) \geq size(ROI_1)$ . This implies that the fractional air content  $F_{2,Air}$  is the fraction of air in the volume defined by the combined set of inhale volume elements that map to the same exhale voxel position. The fraction of air within a single voxel is given simply by:

$$F_{Air}^{voxel} = -\frac{HU_{voxel}}{1000}. \quad (A.13)$$

However, the fraction of air within a volumetric configuration of voxels is rather the total volume of air within the volume of interest (VOI), divided by the total volume of the VOI. Let  $size(ROI_2) = N$ . The total volume of air  $V_{Air}^{VOI}$  is given by the sum of fractional air contents within each voxel, multiplied by the voxel volume:

$$V_{Air}^{VOI} = \sum_{i=1}^N F_{Air}^i V_{voxel} = V_{voxel} \sum_{i=1}^N F_{Air}^i, \quad (A.14)$$

where  $V_{\text{voxel}}$  is the voxel volume. Since the total volume of the VOI is simply  $(NV_{\text{voxel}})$ , we can write an expression for the total fractional air content (i.e., total air volume / total VOI volume):

$$F_{\text{Air}}^{\text{VOI}} = \frac{V_{\text{voxel}} \sum_{i=1}^N F_{\text{Air}}^i}{NV_{\text{voxel}}} = \frac{\sum_{i=1}^N \frac{-HU_i}{1000}}{N} = -\frac{1}{1000} \sum_{i=1}^N HU_i . \quad (\text{A.15})$$

By substitution of equations (A.13) and (A.15) into equation (A.12), we arrive at a formula for calculating the specific ventilation for a single exhale volume element:

$$sVent_{HU} = \frac{\left[ \frac{-1}{1000N} \sum_{i=1}^N HU_i^{\text{inhale}} \right] + \left[ \frac{HU^{\text{exhale}}}{1000} \right]}{\left[ \frac{-HU^{\text{exhale}}}{1000} \right] \left[ 1 + \left( \frac{1}{1000N} \sum_{i=1}^N HU_i^{\text{inhale}} \right) \right]} . \quad (\text{A.16})$$

To obtain the final form of the expression for specific ventilation, we perform additional simplifying operations. Multiplying numerator and denominator by 1000, and simplifying yields:

$$sVent_{HU} = \frac{\left[ -\frac{1}{N} \sum_{i=1}^N HU_i^{\text{inhale}} \right] + HU^{\text{exhale}}}{-HU^{\text{exhale}} \left[ 1 + \left( \frac{1}{1000N} \sum_{i=1}^N HU_i^{\text{inhale}} \right) \right]} . \quad (\text{A.17})$$

Again multiply top and bottom by 1000 and simplify to obtain:

$$sVent_{HU} = 1000 \frac{\left[ -\frac{1}{N} \sum_{i=1}^N HU_i^{\text{inhale}} \right] + HU^{\text{exhale}}}{-HU^{\text{exhale}} \left[ 1000 + \frac{1}{N} \sum_{i=1}^N HU_i^{\text{inhale}} \right]} . \quad (\text{A.18})$$

Multiplying top and bottom by the quantity -1 yields the following:

$$sVent_{HU} = 1000 \frac{\left[ \frac{1}{N} \sum_{i=1}^N HU_i^{inhale} \right] - HU^{exhale}}{HU^{exhale} \left[ 1000 + \frac{1}{N} \sum_{i=1}^N HU_i^{inhale} \right]}. \quad (A.19)$$

Note that the quantity  $\left[ \frac{1}{N} \sum_{i=1}^N HU_i^{inhale} \right]$  is just the average inhale HU within the lung volume that is mapped to the single exhale voxel by DIR spatial transformation. For notational simplicity, let:

$$\bar{H}_{VOI}^{inhale} = \frac{1}{N} \sum_{i=1}^N HU_i^{inhale}. \quad (A.20)$$

Then the final expression for specific ventilation within any exhale lung voxel in terms of registered HU between inhalation and exhalation breathing states is given:

$$sVent_{HU} = \frac{\Delta V}{V_{ex}} = 1000 \frac{\bar{H}_{VOI}^{inhale} - HU^{exhale}}{HU^{exhale} (1000 + \bar{H}_{VOI}^{inhale})}. \quad (A.21)$$

Note that application of equation (A.21) on a voxel-by-voxel basis yields a quantitative map of specific ventilation, inherently co-registered with the exhalation state CT image.

## References

1. Abratt, R.P., P.A. Willcox, and J.A. Smith, *Lung cancer in patients with borderline lung functions--zonal lung perfusion scans at presentation and lung function after high dose irradiation*. Radiother Oncol, 1990. **19**(4): p. 317-22.
2. Petersson, J., Sanchez-Crespo, A., Larsson, S. A., and Mure, M., *Physiological imaging of the lung: single-photon-emission computed tomography (SPECT)*. J Appl Physiol, 2007. **102**(1): p. 468-476.
3. Agnew, J.E., Francis, R.A., Pavia, D., and Clarke, S.W., *Quantitative comparison of 99mTc-aerosol and 81mKr ventilation images*. Clin Phys Physiol Meas, 1982. **3**: p. 21-30.
4. Cabahug, C.J., McPeck, M., Palmer, L.B., Cuccia, A., Atkins, H.L., and Smaldone, G.C., *Utility of technetium-99m-DTPA in determining regional ventilation*. J Nucl Med, 1996. **37**: p. 239-244.
5. Coghe, J., Votion, D., and Lekeux, P., *Comparison between radioactive aerosol, technegas and krypton for ventilation imaging in healthy calves*. Vet J, 2000. **160**: p. 25-32.
6. Simon, B.A., Marcucci, C., Fung, M., and Lele, S. R., *Parameter estimation and confidence intervals for Xe-CT ventilation studies: a Monte Carlo approach*. Journal of Applied Physiology, 1998. **84**(2): p. 709-16.
7. Reinhardt, J.M., Ding, K., Cao, K., Christensen, G. E., Hoffman, E. A., and Bodas, S., *Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation*. Medical Image Analysis, 2008. **12**: p. 752-763.
8. Marcucci, C., Nyhan, D., and Simon, B. A., *Distribution of pulmonary ventilation using Xe-enhanced computed tomography in prone and supine dogs*. Journal of Applied Physiology, 2001. **90**(2): p. 421-30.
9. Gur, D., Drayer, B. P., Borovetz, H. S., Griffith, B. P., Hardesty, R. L., and Wolfson, S. K., *Dynamic computed tomography of the lung: regional ventilation measurements*. Journal of Computer Assisted Tomography, 1979. **3**(6): p. 749-53.
10. Gur, D., Shabason, L., Borovetz, H. S., Herbert, D. L., Reece, G. J., Kennedy, W. H., and Serago, C., *Regional pulmonary ventilation measurements by xenon enhanced dynamic computed tomography: an update*. Journal of Computer Assisted Tomography, 1981. **5**(5): p. 678-83.

11. Edelman, R.R., Hatabu, H., Tadamura, E., Li, W., and Prasad, P. V., *Noninvasive assessment of regional ventilation in the human lung using oxygen-enhanced magnetic resonance imaging*. Nat Med, 1996. **2**(11): p. 1236-9.
12. Kauczor, H.U., Hanke, A., and Van Beek, E. J., *Assessment of lung ventilation by MR imaging: current status and future perspectives*. Eur Radiol, 2002. **12**(8): p. 1962-70.
13. Simon, B.A., *Non-invasive imaging of regional lung function using x-ray computed tomography*. J Clin Monit Comput, 2000. **16**(5-6): p. 433-42.
14. Guerrero, T.M., Sanders, K., Noyola-Martinez, J., Castillo, E., Zhang, Y., Tapia, R., Guerra, R., Borghero, Y., and Komaki, R., *Quantification of regional ventilation from treatment planning CT*. Int J Radiat Oncol Biol Phys, 2005. **62**(3): p. 630 - 634.
15. Guerrero, T., Sanders, K., Castillo, E., Zhang, Y., Bidaut, L., Pan, T., and Komaki, R., *Dynamic ventilation imaging from four-dimensional computed tomography*. Phys Med Biol, 2006. **51**(4): p. 777-91.
16. Yaremko, B.P., Guerrero, T. M., Noyola-Martinez, J., Guerra, R., Lege, D. G., Nguyen, L. T., Balter, P. A., Cox, J. D., and Komaki, R., *Reduction of Normal Lung Irradiation in Locally Advanced Non-Small-Cell Lung Cancer Patients, Using Ventilation Images for Functional Avoidance*. Int J Radiat Oncol Biol Phys 2007. **68**(2): p. 562-571.
17. Dougherty, L., Torigian, D. A., Affusso, J. D., Asmuth, J. C., and Geftter, W. B., *Use of an optical flow method for the analysis of serial CT lung images*. Acad Radiol 2006. **13**(1): p. 14-23.
18. Schnabel, J.A., Tanner, C., Castellano Smith, A. D., Degenhard, A., Leach, M. O., Hose, R., Hill, D. L. G., and Hawkes, D. J., *Validation of non-rigid image registration using finite element methods: application to breast MR images*. IEEE Trans Med Imaging, 2003. **22**(238-247).
19. Uhlemann, F., Morgenstern, U., and Steinmeier, R., *Objective evaluation of three-dimensional image registration algorithms--tools for optimization and evaluation*. Methods Inf Med, 2004. **43**(4): p. 367-70.
20. Guerrero, T., Zhang, G., Huang, T. C., Lin, K. P., *Intrathoracic tumour motion estimation from CT imaging using the 3D optical flow method*. Phys Med Biol, 2004. **49**(17): p. 4147-61.
21. Wang, H., Dong, L., Lu, M.F., Lee, A.L., de Crevoisier, R., Mohan, R., Cox, J.D., Kuban, D.A., and Cheung, R., *Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy*. Int J Radiat Oncol Biol Phys, 2005. **61**(3): p. 725-735.

22. Lu, W., Chen, M. L., Olivera, G. H., Ruchala, K. J., and Mackie, T. R., *Fast free-form deformable registration via calculus of variations*. Phys Med Biol, 2004. **49**(14): p. 3067-87.
23. Kashani, R., Hub, M., Kessler, M. L., and Balter, J. M., *Technical note: a physical phantom for assessment of accuracy of deformable alignment algorithms*. Med Phys, 2007. **34**(7): p. 2785-2788.
24. Kashani, R., Hub, M., Balter, J. M., Kessler, M. L., Dong, L., Zhang, L., Xing, L., Xie, Y., Hawkes, D., Schnabel, J. A., McClelland, J., Joshi, S., Chen, Q., and Lu, W., *Objective assessment of deformable image registration in radiotherapy: A multi-institution study*. Med Phys, 2008. **35**(12): p. 5944-5953.
25. Wang, H., Dong, L., O'Daniel, J., Mohan, R., Garden, A. S., Ang, K. K., Kuban, D. A., Bonnen, M., Change, J. Y., and Cheung, R., *Validation of an accelerated demons algorithm for deformable image registration in radiation therapy*. Phys Med Biol, 2005. **50**(12): p. 2887.
26. Kashani, R., Lam K., Litzenberg, D., and Balter, J., *Technical note: a deformable phantom for dynamic modeling in radiation therapy*. Med Phys, 2007. **34**(1): p. 199-201.
27. Fitzpatrick, J.M., *Detecting Failure, Assessing Success*, in *Medical Image Registration*, M.R. Neuman, Editor. 2001, CRC Press: New York, NY.
28. Wang, H., Garden, A. S., Zhang, L., Wei, X., Ahamad, A., Kuban, D.A., Komaki, R., O'Daniel, J., Zhang, Y., Mohan, R., and Dong, L., *Performance evaluation of automatic anatomy segmentation algorithm on repeat or four-dimensional computed tomography images using deformable image registration method*. Int J Radiat Oncol Biol Phys, 2008. **72**(1): p. 210-219.
29. Kaus, M.R., Brock, K. K., Pekar, V., Dawson, L. A., Nichol, A. M., and Jaffray, D. A., *Assessment of a Model-Based Deformable Image Registration Approach for Radiation Therapy Planning*. Int J Radiat Oncol Biol Phys, 2007. **68**(2): p. 572-580.
30. Boldea, V., Sharp, G. C., Jiang, S. B., and Sarrut, D., *4D-CT lung motion estimation with deformable registration: Quantification of motion nonlinearity and hysteresis*. Med Phys, 2008. **35**(3): p. 1008-1018.
31. Al-Mayah, A., Moseley, J., and Brock, K. K., *Contact surface and material nonlinearity modeling of human lungs*. Phys Med Biol, 2008. **53**(1): p. 305-317.
32. Wu, Z., Rietzel, E., Boldea, V., Sarrut, D., and Sharp, G. C., *Evaluation of deformable registration of patient lung 4DCT with subanatomical region segmentations*. Med Phys, 2008. **35**(2): p. 775-781.

33. Wolthaus, J.W.H., Sonke, J.J., van Herk, M., and Damen, M.F., *Reconstruction of a time-averaged midposition CT scan for radiotherapy planning of lung cancer patients using deformable registration*. Med Phys, 2008. **35**(9): p. 3998-4011.
34. Li, P., Malsch, U., and Bendl, R., *Combination of intensity-based image registration with 3D simulation in radiation therapy*. Phys Med Biol, 2008. **53**: p. 4621-4637.
35. Yang, D., Chaudhari, S. R., Goddu, S. M., Pratt, D., Khullar, D., Deasy, J. O., and El Naqa, I., *Deformable registration of abdominal kilovoltage treatment planning CT and tomotherapy daily megavoltage CT for treatment adaptation*. Med Phys, 2009. **36**(2): p. 329-338.
36. Brock, K.K., Nichol, A.M., Menard, C., Moseley, J.L., Warde, P.R., Catton, C.N., and Jaffray, D.A., *Accuracy and sensitivity of finite element model-based deformable registration of the prostate*. Med Phys, 2008. **35**(9): p. 4019-4025.
37. Lu, W., Olivera, G.H., Chen, Q., Ruchala, K.J., Haimerl, J., Meeks, S.L., Langen, K.M., and Kupelian, P. A., *Deformable registration of the planning image (kVCT) and the daily images (MVCT) for adaptive radiation therapy*. Phys Med Biol, 2006. **51**: p. 4357-4374.
38. Dawood, M., Buther, F., Jiang, X., and Schafers, K.P., *Respiratory motion correction in 3-D PET data with advanced optical flow algorithms*. IEEE Trans Med Imaging, 2008. **27**(8): p. 1164-1175.
39. Wijesooriya, K., Weiss, E., Dill, V., Dong, L., Mohan, R., Joshi, S., and Keall, P. J., *Quantifying the accuracy of automated structure segmentation in 4D CT images using a deformable image registration algorithm*. Medical Physics, 2008. **35**(4): p. 1251-1260.
40. Wirth, M., Fraschini, M., Masek, M., and Bruynooghe, M., *Performance evaluation in image processing*. EURASIP Journal on Applied Signal Processing, 2006. **2006**: p. 1-3.
41. Zhong, H., Peters, T., Siebers, J.V., *FEM-based evaluation of deformable image registration for radiation therapy*. Phys Med Biol, 2007. **52**: p. 4721-4738.
42. Cai, J., Miller, G.W., Altes, T.A., Read, P.W., Benedict, S. H., deLange, E.E., Cates, G. D., Brookeman, J.R., Mugler, J.P., and Sheng, K., *Direct measurement of lung motion using hyperpolarized helium-3 MR tagging*. Int J Radiat Oncol Biol Phys 2007. **68**(3): p. 650-653.
43. Kwa, S.L.S., Theuws, J. C. M., van Herk, M., Muller, S. H., and Lebesque, J. V., *Application of chamfer matching in three-dimensional correlation of CT-SPECT and CT-CT of the lungs*. Radiotherapy and Oncology, 1995. **37**(Supplement): p. S3-S67.
44. Marks, L.B., Spencer, D. P., Sherouse, G. W., Bentel, G., Clough, R., Vann, K., Jaszczak, R., Coleman, R. E., and Prosnitz, L. R., *The role of three dimensional functional lung*

- imaging in radiation treatment planning: the functional dose-volume histogram.* Int J Radiat Oncol Biol Phys, 1995. **33**(1): p. 65-75.
45. Marks, L.B., Munley, M. T., Spencer, D. P., Sherouse, G. W., Bentel, G. C., Hoppenworth, J., Chew, M., Jaszczak, R. J., Coleman, R. E., and Prosnitz, L. R., *Quantification of radiation-induced regional lung injury with perfusion imaging.* Int J Radiat Oncol Biol Phys 1997. **38**(2): p. 399-409.
  46. Fan, M., Marks, L. B., Hollis, D., Bentel, G. G., Anscher, M. S., Sibley, G., Coleman, R. E., Jaszczak, R. J., and Munley, M. T., *Can we predict radiation-induced changes in pulmonary function based on the sum of predicted regional dysfunction?* Journal of Clinical Oncology., 2001. **19**(2): p. 543-50.
  47. Nioutsikou, E., Partridge, M., Bedford, J. L., and Webb, S., *Prediction of radiation-induced normal tissue complications in radiotherapy using functional image data.* Phys Med Biol 2005. **50**(6): p. 1035-1046.
  48. Graham, M.V., Purdy, J. A., Emami, B., Harms, W., Bosch, W., Lockett, M. A., and Perez, C. A., *Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC).* Int J Radiat Oncol Biol Phys, 1999. **45**(2): p. 323-9.
  49. Lee, H.K., Vaporciyan, A. A., Cox, J. D., Tucker, S. L., Putnam, J. B., Jr., Ajani, J. A., Liao, Z., Swisher, S. G., Roth, J. A., Smythe, W. R., Walsh, G. L., Mohan, R., Liu, H. H., Mooring, D., and Komaki, R., *Postoperative pulmonary complications after preoperative chemoradiation for esophageal carcinoma: Correlation with pulmonary dose-volume histogram parameters.* Int J Radiat Oncol Biol Phys, 2003. **57**(5): p. 1317-22.
  50. Lyman, J.T., *Complication probability as assessed from dose-volume histograms.* Radiation Research. Supplement, 1985. **8**: p. S13-9.
  51. Altemeier, W.A., S. McKinney, and R.W. Glenny, *Fractal nature of regional ventilation distribution.* Journal of Applied Physiology, 2000. **88**(5): p. 1551-7.
  52. Glenny, R.W., S.L. Bernard, and H.T. Robertson, *Pulmonary blood flow remains fractal down to the level of gas exchange.* Journal of Applied Physiology, 2000. **89**(2): p. 742-8.
  53. Venegas, J.G. and G.G. Galletti, *Low-pass filtering, a new method of fractal analysis: application to PET images of pulmonary blood flow.* Journal of Applied Physiology, 2000. **88**(4): p. 1365-73.
  54. Gee, J.C. *Performance evaluation of medical image processing algorithms.* in *Medical Imaging 2000: Image Processing.* 2000. San Diego, CA, USA: SPIE.



55. Lehmann, T.M. *From plastic to gold: a unified classification scheme for reference standards in medical image processing*. in *SPIE Medical Imaging*. 2002. San Diego, CA: Society of Photo-Optical Instrumentation Engineers.
56. Rietzel, E., Chen, G. T. Y., *Deformable registration of 4D computed tomography data*. Med Phys, 2006. **33**(11): p. 4423-4430.
57. Brock, K.K., Sharpe, M. B., Dawson, L. A., Kim, S. M., and Jaffray, D. A., *Accuracy of finite element model-based multi-organ deformable image registration*. Med Phys, 2005. **32**(6): p. 1647-1659.
58. Coselmon, M.M., Balter, J. M., McShan, D. L., and Kessler, M. L., *Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines*. Med Phys, 2004. **31**(11): p. 2942-8.
59. Keall, P.J., Joshi, S., Vedam, S. S., Siebers, J. V., Kini, V. R., and Mohan, R., *Four-dimensional radiotherapy planning for DMLC-based respiratory motion tracking*. Med Phys, 2005. **32**(4): p. 942-951.
60. Pevsner, A., Davis, B., Joshi, S., Hertanto, A., Mechalakos, J., Yorke, E., Rosenzweig, K., Nehmeh, S., Erdi, Y. E., Humm, J. L., Larson, S., Ling, C. C., and Mageras, G. S., *Evaluation of an automated deformable image matching method for quantifying lung motion in respiration-correlated CT images*. Med Phys, 2006. **33**(2): p. 369-376.
61. Sarrut, D., Boldea, V., Miguet, S., and Ginestet, C., *Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans*. Med Phys, 2006. **33**(3): p. 605-17.
62. Lewis, J.P., *Fast Template Matching*. Vision Interface, 1995: p. 120-123.
63. Gonzalez, R.C. and R.E. Woods, *Digital Image Processing*. 3 ed. 2008, Upper Saddle River, NJ: Prentice-Hall, Inc. 954.
64. Casella, G. and R. Berger, *Statistical Inference*. 2 ed. 2001, New York, NY: Duxbury Press.
65. Horn, B.K.P. and B.G. Schunck, *Determining optical flow*. Artificial Intelligence, 1981. **17**: p. 185-203.
66. Beauchemin, S.S. and J.L. Barron, *The computation of optical flow*. ACM Computing Surveys, 1995. **27**(5): p. 433 - 466.
67. Press, W.H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., *Numerical Recipes in C++: The art of scientific computing*. 2002, Cambridge: Cambridge University Press.

68. Hu, S., E.A. Hoffman, and J.M. Reinhardt, *Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images*. IEEE Trans Med Imaging, 2001. **20**(6): p. 490-8.
69. Schaefer, S., T. McPhail, and J. Warren, *Image deformation using moving least squares*, in *ACM SIGGRAPH 2006 Papers*. 2006, ACM Press: Boston, MA.
70. Sharpe, M. and K.K. Brock, *Quality assurance of serial 3D image registration, fusion, and segmentation*. Int J Radiat Oncol Biol Phys, 2008. **71**(1S): p. S33-S37.
71. Bevington, P.R. and D.K. Robinson, *Data Reduction and Error Analysis*. Vol. 3. 2003, New York, NY: McGraw Hill.
72. Sarrut, D., Delhay, S., Villard, P. F., Boldea V. A., Beuve M. A., and Clarysse P. A., *A Comparison Framework for Breathing Motion Estimation Methods From 4-D Imaging*. IEEE Trans Med Imaging, 2007. **26**(12): p. 1636-1648.
73. Alderson, P.O. and B.R. Line, *Scintigraphic evaluation of regional pulmonary ventilation*. Semin Nucl Med, 1980. **10**(3): p. 219-42.
74. Burch, W.M., P.J. Sullivan, and C.J. McLaren, *Technegas--a new ventilation agent for lung scanning*. Nucl Med Commun, 1986. **7**(12): p. 865-71.
75. Suga, K., *Technical and analytical advances in pulmonary ventilation SPECT with xenon-133 gas and Tc-99m-Technegas*. Ann Nucl Med, 2002. **16**(5): p. 303-10.
76. Sullivan, P.J., Burke, W. M., Burch, W. M., and Lomas, F. E., *A clinical comparison of Technegas and xenon-133 in 50 patients with suspected pulmonary embolus*. Chest, 1988. **94**(2): p. 300-4.
77. Schuster, D.P., *Positron emission tomography: theory and its application to the study of lung disease*. American Review of Respiratory Disease., 1989. **139**(3): p. 818-40.
78. Rhodes, C.G., Valind, S. O., Brudin, L. H., Wollmer, P. E., Jones, T., Buckingham, P. D., and Hughes, J. M., *Quantification of regional v/q ratios in humans by use of pet. Procedure and normal values*. J Appl Physiol, 1989. **66**(4): p. 1905-1913.
79. Mijailovich, S.M., S. Treppo, and J.G. Venegas, *Effects of lung motion and tracer kinetics corrections on PET imaging of pulmonary function*. Journal of Applied Physiology., 1997. **82**(4): p. 1154-62.
80. Richard, J.-C., Janier, M., Lavenne, F., Tourvieille, C., Le Bars, D., Costes, N., Gimenez, G., and Guerin, C., *Quantitative assessment of regional alveolar ventilation and gas volume using 13n-n2 washout and pet*. J Nucl Med, 2005. **46**(8): p. 1375-1383.

81. Wellman, T.J., Winkler, T., Costa, E. L. V., Musch, G., Harris, R. S., Venegas, J. G., and Melo, M. F. V., *Measurement of regional specific lung volume change using respiratory-gated pet of inhaled  $^{13}\text{N}$ -nitrogen*. J Nucl Med, 2010. **51**(4): p. 646-653.
82. Valind, S.O., Rhodes, C. G., Brudin, L. H., and Jones, T., *Measurements of regional ventilation pulmonary gas volume: Theory and error analysis with special reference to positron emission tomography*. J Nucl Med, 1991. **32**(10): p. 1937-1944.
83. Albert, M.S., Cates, G. D., Driehuys, B., Happer, W., Saam, B., Springer, C. S., Jr., and Wishnia, A., *Biological magnetic resonance imaging using laser-polarized  $^{129}\text{Xe}$* . Nature, 1994. **370**(6486): p. 199-201.
84. Endo, M., Tsunoo, T., Kandatsu, S., Tanada, S., Aradate, H., and Saito, Y., *Four-dimensional computed tomography (4D CT)--concepts and preliminary development*. Radiat Med, 2003. **21**(1): p. 17-22.
85. Vedam, S.S., Keall, P. J., Kini, V. R., Mostafavi, H., Shukla, H. P., and Mohan, R., *Acquiring a four-dimensional computed tomography dataset using an external respiratory signal*. Phys Med Biol, 2003. **48**(1): p. 45-62.
86. Keall, P.J., Starkschall, G., Shukla, H., Forster, K. M., Ortiz, V., Stevens, C. W., Vedam, S. S., George, R., Guerrero, T., and Mohan, R., *Acquiring 4D thoracic CT scans using a multislice helical method*. Physics in Medicine & Biology, 2004. **49**(10): p. 2053-67.
87. Nehmeh, S.A., Erdi, Y. E., Pan, T., Yorke, E., Mageras, G. S., Rosenzweig, K. E., Schoder, H., Mostafavi, H., Squire, O., Pevsner, A., Larson, S. M., and Humm, J. L., *Quantitation of respiratory motion during 4D-PET/CT acquisition*. Med Phys, 2004. **31**(6): p. 1333-8.
88. Underberg, R.W., Lagerwaard, F. J., Cuijpers, J. P., Slotman, B. J., van Sornsens de Koste, J. R., and Senan, S., *Four-dimensional CT scans for treatment planning in stereotactic radiotherapy for stage I lung cancer*. Int J Radiat Oncol Biol Phys, 2004. **60**(4): p. 1283-90.
89. Lu, W., Parikh, P. J., El Naqa, I. M., Nystrom, M. M., Hubenschmidt, J. P., Wahab, S. H., Mutic, S., Singh, A. K., Christensen, G. E., Bradley, J. D., and Low, D. A., *Quantitation of the reconstruction quality of a four-dimensional computed tomography process for lung cancer patients*. Med Phys 2005. **32**(4): p. 890-901.
90. Li, T., Schreibmann, E., Thorndyke, B., Tillman, G., Boyer, A., Koong, A., Goodman, K., and Xing, L., *Radiation dose reduction in four-dimensional computed tomography*. Med Phys, 2005. **32**(12): p. 3650-60.
91. Fleming, E.H., *Functions of Several Variables*. 1977, New York: Springer-Verlag.

92. Guerrero, T.M., Castillo, R., Sanders, K., Price, R., Komaki, R., and Cody, D. D., *Novel method to calculate pulmonary compliance images in rodents from computed tomography acquired at constant pressures*. Phys Med Biol, 2006. **51**(5): p. 1101-1112.
93. Pan, T., Lee, T. Y., Rietzel, E., and Chen, G. T., *4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT*. Med Phys, 2004. **31**(2): p. 333-40.
94. Shioyama, Y., Jang, S.Y., Liu, H.H., Guerrero, T., Wang, X., Gayed, I.W., Erwin, W. D., Liao, Z., Chang, J.Y., Jeter, M., Yaremko, B.P., Borghero, Y.O., Cox, J.B., Komaki, R., and Mohan, R., *Preserving functional lung using perfusion imaging and intensity-modulated radiation therapy for advanced-stage non-small cell lung cancer*. Int. J. Radiat. Oncol. Biol. Phys. , 2007. **68**(5): p. 1349-1358.
95. Ireland, R.H., Bragg, C.M., McJury, M., Woodhouse, N., Fichele, S., van Beek, E. J. R., Wild, J. M., and Hatton, M.Q., *Feasibility of Image Registration and Intensity-Modulated Radiotherapy Planning With Hyperpolarized Helium-3 Magnetic Resonance Imaging for Non-Small-Cell Lung Cancer*. Int J Radiat Oncol Biol Phys 2007. **68**(1): p. 273-281.
96. Hodge, C.W., Tome, W. A., Fain, S. B., Bentzen, S. M., and Mehta, M. P., *On the use of hyperpolarized helium MRI for conformal avoidance lung radiotherapy*. Med Dosim, 2010. **35**(4): p. 297-303.
97. Munawar, I., Yaremko, B. P., Craig, J., Oliver, M., Gaede, S., Rodrigues, G., Yu, E., Henderson Reid, R., Leung, E., Urbain, J. L., Chen, J., and Wong, E., *Intensity modulated radiotherapy of non-small-cell lung cancer incorporating SPECT ventilation imaging*. Med Phys, 2010. **37**(4): p. 1863-1872.
98. Castillo, E., Castillo, R., Martinez, J., Shenoy, M., and Guerrero, T., *Four-dimensional deformable image registration using trajectory modeling*. Phys Med Biol, 2010. **55**: p. 305-327.
99. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., and Guerrero, T., *A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets*. Phys Med Biol, 2009. **54**: p. 1849-1870.
100. Guerrero, T., Castillo, R., Noyola-Martinez, J., Torres, M., Zhou, X., Guerra, R., Cody, D., Komaki, R., and Travis, E., *Reduction of pulmonary compliance found with high-resolution computed tomography in irradiated mice*. Int J Radiat Oncol Biol Phys 2007. **67**(3): p. 879-887.
101. Nocedal, J. and S.J. Wright, *Numerical Optimization*. Springer Series in Operations Research. 1999, New York: Springer.

102. Dice, R., *Measure of the amount of ecological association between species*. Ecology, 1945. **26**(2): p. 297-302.
103. Riegel, A., Chang, J. Y., Vedam, S. S., Johnson, V., Chi, M. P. C., and Pan, T., *Cine computed tomography without respiratory surrogate in planning stereotactic radiotherapy for non-small-cell lung cancer*. Int J Radiat Oncol Biol Phys, 2009. **73**(2): p. 433-441.
104. Fuld, M.K., Easley, R. B., Saba, O. I., Chon, D., Reinhardt, J. M., Hoffman, E. A., and Simon, B. A., *CT-measured regional specific volume change reflects regional ventilation in supine sheep*. J Appl Physiol, 2008. **104**(4): p. 1177-84.
105. Kabus, S., von Berg, J., Yamamoto, T., Opfer, R., and Keall, P. J., *Lung ventilation estimation based on 4D-CT imaging*. Proc of the First International Workshop on Pulmonary Image Analysis, MICCAI, 2008: p. 73-81.
106. Kabus, S., et al. *Lung ventilation estimation based on 4D-CT imaging*. in *First International Workshop on Pulmonary Image Analysis*. 2008: Springer: New York.
107. Bates, E.L., Bragg, C. M., Wild, J. M., Hatton, M. Q. F., and Ireland, R. H., *Functional image-based radiotherapy planning for non-small-cell lung cancer: A simulation study*. Radiother Oncol, 2009. **93**(1): p. 32-36.
108. Marks, L.B., *Physiology-based studies of radiation-induced normal tissue injury*. Radiotherapy & Oncology., 1999. **51**(2): p. 101-3.
109. Theuws, J.C.M., Kwa, S. L. S., Wagenaar, A. C., Boersmaa, L. J., Damena, E. M. F., Mullera, S. H., Baasc, P., and Lebesque, J. V., *Dose-effect relations for early local pulmonary injury after irradiation for malignant lymphoma and breast cancer*. Radiotherapy and Oncology, 1998. **48**(1): p. 33-43.
110. Narabayashi, I., T. Suematsu, and C. Suematsu, *Assessment of SPECT ventilation-perfusion imaging in patients with lung cancer*. European Journal of Nuclear Medicine and Molecular Imaging, 1986. **12**(7): p. 329-332.
111. Suga, K., Kawakami, Y., Zaki, M., Yamashita, T., Shimizu, K., and Matsunaga, N., *Clinical utility of co-registered respiratory-gated <sup>99m</sup>Tc-Technegas/MAA SPECT-C images in the assessment of regional lung functional impairment in patients with lung cancer*. Eur J Nucl Med Mol Imaging, 2004. **31**(9): p. 1280-1290.
112. Castillo, R., Castillo, E., Martinez, J., and Guerrero, T., *Ventilation from four-dimensional computed tomography: density versus Jacobian methods*. Phys Med Biol, 2010. **55**(16): p. 4661-4685.

113. Reinhardt, J., Christensen, G., Hoffman, E., Ding, K., and Cao, K., *Registration-Derived Estimates of Local Lung Expansion as Surrogates for Regional Ventilation*, in *Information Processing in Medical Imaging*. 2007. p. 763-774.
114. Yamamoto, T., Kabus, S., von Berg, J., Lorenz, C., and Keall, P. J., *Impact of four-dimensional computed tomography pulmonary ventilation imaging-based functional avoidance for lung cancer radiotherapy*. *Int J Radiat Oncol Biol Phys*, 2010. **79**(1): p. 279-88.
115. Castillo, E., Castillo, R., Gu, X., Martinez, J., Fuentes, D., Jiang, S., Friedman, P., and Guerrero, T., *Deformable image registration for breath-hold CT image pairs from the COPDgene study*. 2011. **Submitted: Medical Image Analysis**.
116. Ezhil, M., Starkschall, G., Mohan, R., Cox, J., and Komaki, R., *Validation of a model-based segmentation approach to propagating normal anatomic regions of interest through the 10 phases of respiration*. *Int J Radiat Oncol Biol Phys*, 2008. **71**(3): p. 900-906.
117. Hochberg, Y. and A.C. Tamhane, *Multiple comparison procedures*. 1987: John Wiley & Sons, Inc.
118. Zou, K.H., Warfield, S. K., Bharatha, A., Tempany, C. M. C., Kaus, M. R., Haker, S. J., Wells, W. M., Jolesz, F. A., and Kikinis, R., *Statistical validation of image segmentation quality based on a spatial overlap index*. *Acad Radiol*, 2004. **11**(2): p. 178-189.
119. Bartko, J.J., *Measurement and reliability: statistical thinking considerations*. *Schizophr Bull*, 1991. **17**(3): p. 483-489.
120. Yamamoto, T., Kabus, S., von Berg, J., Lorenz, C., Goris, M. L., Loo, B. W., and Keall P. J., *Evaluation of four-dimensional computed tomography pulmonary ventilation imaging by comparison with single photon emission computed tomography scans for a lung cancer patient*, in *Third international workshop on pulmonary image analysis*. 2010. p. 117-128.
121. Yamamoto, T., Langner, U., Loo, B. W., Shen, J., and Keall, P. J., *Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients*. *Int J Radiat Oncol Biol Phys*, 2008. **72**(4): p. 1250-1258.
122. Yin, L., Shcherbinin, S., Celler, A., Thompson, A., Fua, T. F., Liu, M., Duzenli, C., Gill, B., Sheehan, F., Powe, J., Worsley, D., Marks, L., and Moiseenko, V., *Incorporating quantitative single photon emission computed tomography into radiation therapy treatment planning for lung cancer: impact of attenuation and scatter correction on the*

- single photon emission computed tomography-weighted mean dose and functional lung segmentation.* Int J Radiat Oncol Biol Phys, 2010. **78**(2): p. 587-94.
123. Cournand, A., *Pulmonary circulation; its control in man, with some remarks on methodology.* Science, 1957. **125**(3260): p. 1231-5.
  124. Brower, R., Wise, R. A., Hassapoyannes, C., Bromberger-Barnea, B., and Permutt, S., *Effect of lung inflation on lung blood volume and pulmonary venous flow.* J Appl Physiol, 1985. **58**(3): p. 954-63.
  125. Castillo, E., Castillo, R., Zhang, Y., and Guerrero, T., *Compressible image registration for thoracic computed tomography images.* Journal of Medical and Biological Engineering, 2009. **29**: p. 222-233.
  126. Marks, L.B., Spencer, D. P., Bentel, G. C., Ray, S. K., Sherouse, G. W., Sontag, M. R., Coleman, R. E., Jaszczak, R. J., Turkington, T. G., and Tapson, V., *The utility of SPECT lung perfusion scans in minimizing and assessing the physiologic consequences of thoracic irradiation.* Int J Radiat Oncol Biol Phys, 1993. **26**(4): p. 659-68.
  127. Marks, L.B., Munley, M. T., Bentel, G. C., Zhou, S. M., Hollis, D., Scarfone, C., Sibley, G. S., Kong, F. M., Jirtle, R., Jaszczak, R., Coleman, R. E., Tapson, V., and Anscher, M., *Physical and biological predictors of changes in whole-lung function following thoracic irradiation.* International Journal of Radiation Oncology, Biology, Physics., 1997. **39**(3): p. 563-70.
  128. Garipagaoglu, M., Munley, M. T., Hollis, D., Poulson, J. M., Bentel, G. C., Sibley, G., Anscher, M. S., Fan, M., Jaszczak, R. J., Coleman, R. E., and Marks, L. B., *The effect of patient-specific factors on radiation-induced regional lung injury.* Int J Radiat Oncol Biol Phys 1999. **45**(2): p. 331-8.
  129. Gu, X., Pan, H., Liang, Y., Castillo, R., Yang, D., Choi, D., Castillo, E., Majumdar, A., Guerrero, T., and Jiang, S. B., *Implementation and evaluation of various demons deformable image registration algorithms on a GPU.* Phys Med Biol, 2010. **55**: p. 207-219.
  130. Coxson, H.O., Mayo, J., Lam, S., Santyr, G., Parraga, G., and Sin, D. D., *New and Current Clinical Imaging Techniques to Study Chronic Obstructive Pulmonary Disease.* Am J Respir Crit Care Med, 2009. **180**(7): p. 588-597.
  131. Bender, E.T. and W.A. Tome, *The utilization of consistency metrics for error analysis in deformable image registration.* Phys Med Biol, 2009. **54**: p. 5561-5577.

## **Vita**

Richard Castillo was born in San Antonio, Texas on November 10<sup>th</sup>, 1981, the son of Edward and Carmen Castillo. As an undergraduate he attended Trinity University in San Antonio, where he earned the degree of Bachelor of Science with a double major in Physics and Mathematics in 2004. In August 2004, he entered The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences. In December 2007 he earned the degree of Master of Science in Medical Physics.

Permanent address:

3203 Heather Hill

San Antonio, TX 78228