QUANTITATIVE ASSESSMENT OF VOLUME CHANGE IN TUMORS USING IMAGE REGISTRATION

by

Saradwata Sarkar

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in the University of Michigan
2011

Doctoral Committee:

Professor Charles R. Meyer, Chair
Professor Jeffrey A. Fessler
Professor Douglas C. Noll
Associate Professor Marc L. Kessler

To Baba and Ma, hope this makes you proud.

Acknowledgements

The time I have spent in Michigan has been one of the happiest in my life. Ann Arbor is a lovely place in itself but it is the people here that truly made this experience special. I take this opportunity to thank some of the people without whom this thesis would not have been possible.

I owe my deepest gratitude to my advisor Professor Charles Meyer for his unstinting support, encouragement and advice through these years. I am glad I knocked on his door five years ago to ask if he had an opening in his lab. What I have learned from him over these years has made this experience truly priceless. I am grateful to Professor Noll for supporting me with a GSI position through a difficult period and for teaching me BME 516 which introduced me to biomedical imaging. The opportunities I have had to interact with Professor Fessler have been one of the highlights of my academic experience at Michigan and also a great honor. I would like to thank Professor Kessler for his enthusiasm and encouragement and Professor Johnson for helping me with statistics.

I would also like to thank Bing for being a great mentor and friend, Peyton for solving innumerable different problems that only he could have solved, Gary for his excellent technical support and Dr. Kim for her kind encouragement. To my DIPL friends present and past, Ram, Roshni, Desmond, Jin, Hiro, Rahul, my deepest thanks. I would also like

to thank Maria Steele in BME for all her help through these years. She is one of the most helpful and kindest people that I know.

Now for my crazy and awesome bunch of friends – ‘paunchy’ Shiva ‘anna’, ‘look at that road!’ Shyam, ‘one sec’ Choppa, ‘stf’ Bala, ‘scene boy’ Manix, ‘kungfu panda’ Divya, ‘ah! let me see’ Dinki, ‘so funny’ Shreya, ‘iomegabasha’ Ram, ‘jignesh’ Smpa and ‘old man’ Rao – thank you and I love you all. I also thank Suchi, Meki, Aarthi, Vijay, Neeti, Shakti, Anu, Pinak, Pratik, Dinesh, dribbletrouble and quizclub members and friends from my skit groups.

A big thank you to my didi, the greatest sister in the world and my sweetest grandmother, Thakuma. A special thanks to my little niece Anusha and littler nephew Anurag and my wonderful brother-in-law Arghyada. Finally and most importantly I would like to thank my parents without whom none of this would have been possible. Baba and Ma, this has been as much your dream as it has been mine. I hope this makes you proud.

This work was funded by DHHS NIH NCI grants 1P01CA87634 and 1P01CA85878.

Table of Contents

List of Figures

Figure

List of Tables

Table

Abstract

QUANTITATIVE ASSESSMENT OF VOLUME CHANGE IN TUMORS USING IMAGE REGISTRATION

by

Saradwata Sarkar

Chair: Charles R. Meyer

Tumor volume change is currently under investigation by numerous groups including the Radiological Society of North America's Quantitative Imaging Biomarker Alliance as a biomarker for assessing response to therapy. Accurate early quantification of tumor volume change could lead to shorter phase III clinical trials as well as the possibility of interactively adapting an individual patient's therapy such as drug or dose modification to achieve optimal response.

The temporal evolution of tumor mass in a patient is a complicated process in three spatial dimensions. Standard clinical techniques for tumor response evaluation, like the unidimensional Response Evaluation Criteria In Solid Tumors 1.1 (RECIST 1.1) or

bidimensional World Health Organization (WHO) measurement methods have limited effectiveness in accurately assessing small, early changes. Most current approaches estimate tumor volume change indirectly by individually segmenting the interval exams in 3D and then subtracting the volumes between time points to obtain a change estimate. Such indirect methods tend to ignore the crucial mutual information present in the tumors. Ensuring the consistency of these independent segmentations across intervals can hence become a significant challenge.

This thesis develops a low noise, low bias direct algorithm to measure volume change using 3D image registration. Tumor pairs are spatially registered across interval exams and volumetric change is calculated by summing local scale changes obtained from the Jacobian map of the deformation. Such an approach not only provides a summary statistic of volume change but can also potentially show regions of differential growth and contraction across the lesion that are difficult to obtain via binary segmentation approaches. The registration-based algorithm is evaluated using synthetic and in vivo interval scans where true tumor volume change is unequivocally known. The 95% confidence error interval of measured volume change was (-8.93%, 10.49%) and (-7.69%, 8.83%) using mutual information and normalized cross correlation, respectively, as similarity measures for registration. To the best of this author's knowledge, these are the tightest bounds reported thus far for zero-change *in vivo* studies. No statistically significant evidence of functional bias was found for the registration-based volume change measurement algorithm.

Statistical models are developed to show that using the registration-based algorithm the error in measuring volume change increases with increase in tumor volume and decreases with the increase in tumor's normalized mutual information, even when that is not the similarity measure being optimized. The developed registration-based change measurement algorithm is also compared with other approaches to demonstrate that it has the potential to outperform indirect segmentation-based volume change measurement methods. The potential of an accurate registration-based volume change measurement algorithm in tracking progression of chronic obstructive pulmonary disease is also suggested through an initial study on a normal and a diseased patient.

Chapter 1

Introduction

1.1 Problem Introduction and Motivation

The response to therapy of tumors often serves as a valuable surrogate marker for predicting clinical outcome in a patient. As far back as the early 1960s, an attempt was made to study the response of tumors to therapeutic agents [1]. The 1979 World Health Organization (WHO) handbook for reporting results of cancer treatment [2], [3] was the first standardized process to unify tumor size change measurement and tumor response to therapy. The WHO criterion for size measurement involves a bilinear measurement of the greatest diameter and the longest perpendicular diameter in each lesion. The sum of the product of the two diameters of all lesions is used to classify a patient into one of four categories – complete response, partial response, stable disease and progressive disease. In 2000, Therasse et al. [4] introduced the Response Evaluation Criteria In Solid Tumors (RECIST) guidelines which employ a unidimensional measurement technique (sum of the longest diameters of target lesions in the axial plane) to quantify tumor burden. The simpler 1D RECIST soon became more popular than 2D WHO [5], [6], [7] and also the most widely used criteria in cancer clinical trials [8]. Today standard clinical radiological

techniques for change assessment largely depend on the RECIST or rather the newer RECIST 1.1 [9] and to a somewhat lesser extent, the WHO criteria.

The temporal evolution of tumor mass in response to cytotoxic therapy is, however, a complicated process in three spatial dimensions. Standard clinical techniques for tumor response evaluation, like the unidimensional RECIST 1.1 or bidimensional WHO measurement methods make many assumptions [10], [11], [12] and may be hopelessly challenged in accurately assessing small, early changes of irregularly shaped lesions. These criteria work best when the tumors under consideration are spherical and their change of shape due to progression of disease or response to therapy is symmetrical. However, tumors often exhibit complicated geometry and their shape evolution is not typically symmetric either. Also, making diameter measurements only along the transverse plane ignores changes to the lesion along the scanner's axis.

With the advancement in three dimensional image acquisitions, the value of volumetric measurement of tumors for response assessment has now come to the fore. Volume change measurement is likely a more reliable marker for response assessment. Studies by Partridge et al. [13] and Yeo et al. [14] have suggested that tumor volume change measurements are more sensitive indicators of pathologic tumor response than tumor diameter measurements. Levine et al. [15] found that volumetric CT was dramatically better in size determination than RECIST especially for small lesions. There are reasons to believe that accurate early quantification of tumor volume change could lead to shorter phase III clinical trials as well as the possibility of interactively adapting an individual patient's therapy such as drug or dose modification to achieve optimal response [16],

[17]. As such, volume change quantification is now a major topic of interest for the NIH, the FDA and pharmaceutical companies [18].

However, until the present time, finding accurate techniques for volumetric change measurement of tumors remains a significant challenge. One approach to solve the problem is to individually segment the interval exams in 3D and subtract the volumes between time points for a change estimate [19], [20]. This is an indirect method in that volume change is not measured directly and the efficacy of these techniques will depend on the accuracy and consistency of the segmentation across interval scans [21]. This thesis investigates an alternative, less explored but direct approach to volume change quantification using image registration. The goal is to develop low noise estimators of volume change that may be able to detect statistically smaller effect sizes, i.e., smaller tumor volume changes that can be observed in shorter intervals. The following section provides an overview of image registration.

1.2 Image Registration Overview

Often imaging applications require the acquisition and analysis of multiple views or images of the same scene. The multiple images might be acquired from different viewpoints or at different times or using different sensors [22]. The goal of image registration is to determine a geometric transformation between two or more such corresponding images in order to compare or combine the information jointly provided by them. Today image registration is widely used in the fields of remote sensing, astrophotography, cartography, computer vision and medical imaging to name just a few. In the field of medical imaging, image registration has been successfully used for

applications like treatment monitoring, patient motion correction, construction of probabilistic population atlases and combining complementary information from multiple modalities (e.g. from CT and MRI) for diagnostic purposes.

The basic notion of image registration, as stated above, is to find a geometric transformation that will make a homologous or floating image of an object *similar* to a reference or source image of the object [23]. To state this formally, we seek to estimate $\hat{T}$ such that

$$\hat{T} = \ argmax_T(S(\ U(x_U), V^T(x_U)))$$

Here U and V denote respectively the reference and homologous images of the object. They can be thought of as mappings of coordinate locations x in their domain or field of view Ω to intensity values, as stated below.

$$U: x_U \in \ \Omega_U \ \mapsto U(x_U)$$

$$V: x_V \in \ \Omega_V \ \mapsto V(x_V)$$

T is a spatial transformation that maps the coordinate frames of the two images U and V such that $U(x_U)$ and $V^T(x_U)$ represent the same location post registration, i.e.,

$$T: x_U \mapsto x_V \ \Leftrightarrow T(x_U) = \ x_V$$

It should be noted that due to the discrete nature of the image domains Ω_U and Ω_V the T we refer to is not purely a spatial transformation but also has to take sampling and interpolation issues into account. S is a similarity measure for the two images calculated on the overlap domain $\Omega_{U,V}^T$ of the two images where

$$\Omega_{U,V}^{T} = \{x_U \in \Omega_U \,|T^{-1}(x_U) \in \Omega_V\}$$

Registration algorithms usually iteratively estimate the transformation $\hat{T}$ to optimize a suitable similarity measure or cost function S.

The formal statement above illuminates the key ingredients of any registration algorithm:

- Spatial Transformations
- Sampling and Interpolation
- Similarity Measures
- Iterative Optimization

Let us look at these in a little more detail.

1.2.1 Spatial Transformations

These are geometrical mappings that determine correspondences between the reference and homologous images (usually 3D to 3D) and can be broadly classified into two categories: Linear transformations (Rigid-body and Affine transformations) and Non-linear transformations.

1.2.1.1 Linear transformations

Rigid-body and Affine transformations fall under the category of linear transformations. Rigid-body transformations preserve distances between all points and in essence represent a simple mapping between two Cartesian system of coordinates. For 3D to 3D transformations, the mapping is governed by six degrees of freedom – three translations and three rotations. Rigid-body transformations have often been successfully used for some registration applications involving the brain or other rigid anatomical structures like

individual bones [23]. Affine transformations preserve parallel lines. To the six degrees of freedom of a rigid-body transformation, six more are added – three to allow anisotropic scaling and three to allow skews. Example applications of affine registration in medical imaging are mentioned in [23].

1.2.1.2 Non-linear transformations

These are complex high degree of freedom transformations that map straight lines to curves. An excellent review of geometric transformations used for non-linear registration is provided by Holden [24]. Non-linear transformations can broadly be classified under two categories – transformations derived from physical models like those based on linear elasticity or fluid flow, and transformations that model the deformation using basis functions like radial basis functions, B-splines etc.

Linear elasticity based models are founded on Hooke's law of linear relationship between stress and strain tensors and solved by the Navier-Cauchy linear elastic PDE. Fluid flow based models are governed by the Navier-Stokes equation applied to velocity fields. For details, the reader is referred to [25], [26].

Thin plate splines (TPS) are perhaps the most commonly used radial basis functions (RBFs). The TPS displacement in 3D, $F(x_U)$, is given below.

$$F(x_U) = Ax_U + B + I\sum_{i=1}^{N} W_i r_i$$

Here A and B define an affine transformation and I is the identity matrix. W_i are the warping coefficients, N are the number of control points or landmarks and r_i are distance functions $\|x_U - x_{U_i}\|$ between the interpolation point x_U and landmark or control point

centre x_{U_i}. A, B and the TPS coefficients W_i are determined by solving the above set of linear equations at the N control point locations in the reference image U [27]. The TPS interpolant minimizes bending energy which functions as an internal regularizer. Large deformations can lead to singularities in the matrix leading to folding without additional regularization. Also, TPS are globally supported and hence computationally complex. There are other topology preserving [24], computationally efficient RBFs with local support that can be used to model deformations - most notably Wu's [28] and Wendland's [29] compactly supported RBFs. These functions, unlike TPS, do not require the polynomial affine part and the coefficients $\{W_i\}$ are solved as

$$KW = F(x_U)$$

where K is a $N \times N$ matrix of radial basis function expansions with

$$K_{ij} = \psi\left(\left\|x_{U_i} - x_{U_j}\right\|\right)$$

Some of Wendland's $\psi_{d,k}$ functions (positive definite in $\mathbb{R}^d$ and C^{2k} continuous) are listed below.

$$\psi_{3,0} = (1-r)_+^2$$

$$\psi_{3,1} = (1-r)_+^4(4r+1)$$

$$\psi_{3,2} = (1-r)_+^6\left({}^{35}/_{3}\, r^2 + 6r + 1\right)$$

where

$$(1-r)_+ = \begin{cases} (1-r), & 0 \le r \le 1 \\ 0, & otherwise \end{cases}$$

The extent of local support a is controlled by a scaling factor on r

$$\psi_a = \psi(r/a)$$

Wendland's functions were suggested by Fornefett for registration [30]. Another popular basis function expansion is cubic B-splines, piecewise polynomials of degree 3 with compact support and C^2 continuous. They are often described by a free-form deformation model as proposed by Rueckert [31].

1.2.2 Sampling and Interpolation

To represent the homologous image in the coordinate frame of the reference, we need to determine not only the corresponding spatial coordinates of the reference image in the homologous image but also the intensity of the homologous image at those coordinates. As mentioned previously, the image domains Ω_U and Ω_V are discrete in nature and intensity values for both images are only defined at the sampled locations, i.e., the coordinates of the two images. Usually an estimated transformation $\hat{T}$ will map coordinates in the reference image domain to non-sampled locations in the homologous image domain where the corresponding intensity values are not known. Interpolation is necessary to determine the intensity value of the homologous image at these coordinates. This interpolation can be implemented by convolving the homologous image with an appropriate kernel. The *sinc* function is ideal for interpolation but impractical due to its infinite extent. In practice interpolation is realized via interpolants of bounded support.

Nearest neighbor interpolation assigns the intensity of the spatially closest neighbor to the coordinate of interest. This makes the interpolation fast but also noisy. Linear interpolation is most popularly used [32]. The intensity at a given coordinate is

determined by a weighted average of the intensities of its neighbors. The weights are determined by the distance of the neighbors from the point of interest. For smoother changes to the joint histogram, Collignon [33] introduced partial volume interpolation. Weights are determined just like in linear interpolation but are not used to interpolate intensities. Instead they are used for fractional updates to the joint histogram at the point of interest and its neighbors. Artifacts related to image interpolation have been studied widely and some of them are discussed in [34].

1.2.3 Similarity Measures

Similarity measures are objective functions that determine how 'close' or 'similar' two images are to each other. The simplest measure is the sum of squared differences (SSD) cost function. It is defined as

$$SSD(U,V) = \frac{1}{N} \sum_{x_U \in \Omega_{U,V}^T} |U(x_U) - V^T(x_U)|^2$$

where N is the number of voxels in the overlap domain of the two images. More robust to the presence of outliers is the sum of absolute differences (SAD) cost function, defined as

$$SAD(U,V) = \frac{1}{N} \sum_{x_U \in \Omega_{U,V}^T} |U(x_U) - V^T(x_U)|$$

When a linear relationship between the intensities of two intra-modality images exists, a common similarity measure employed is normalized cross correlation. Normalized cross correlation is given as

$$NCC(U,V) = \frac{\sum_u (U(x_u) - \bar{U})(V^T(x_u) - \bar{V})}{\sqrt[2]{\{\sum_u (U(x_u) - \bar{U})^2 \sum_u (V^T(x_u) - \bar{V})^2\}}}$$

Here $\bar{U}$ and $\bar{V}$ are mean intensity values of the images U and V^T respectively. Registration is achieved by maximizing the normalized cross correlation of the two images. None of these measures, however, are robust to contrast enhanced differences that are frequently encountered in medical images.

Information theoretic measures are much more robust than the above mentioned measures. Among information theoretic measures, mutual information and its variants are the most widely used for image registration, especially for multimodality image registration. Registration by maximizing mutual information was introduced by Viola [35] and evidence of its versatility can be found in [36]. Mutual Information is computed as

$$MI(U,V) = \sum_u \sum_v p_{UV}^T(u,v) log\left(\frac{p_{UV}^T(u,v)}{p_U(u)p_V^T(v)}\right)$$

where $p_{UV}^T(u,v)$ is the joint density of scans U and V and $p_U(u)$ is the marginal density of U and $p_V^T(v)$ is the marginal of V. The densities are computed via a joint histogram of the images in the overlap domain and are hence dependent upon the transformation.

Mutual information is defined as the amount by which the uncertainty in an image decreases by knowing the other image [32]. In other words,

$$MI(U,V) = H(U) - H(U|V) = H(V) - H(V|U)$$

Here $H(U)$ and $H(V)$ are the Shannon entropies of image *U* and *V* respectively and $H(U|V)$, $H(V|U)$ are conditional entropies. Mutual information is also defined as

$$MI(U,V) = H(U) + H(V) - H(U,V)$$

where $H(U,V)$ is the joint entropy of the two images. Registration of two images can also proceed via the iterative minimization of joint entropy. However, mutual information is much less sensitive to overlap than joint entropy and hence usually preferred.

Studholme proposed a normalized measure of mutual information and found it to be less sensitive to changes in overlap than mutual information [37]. Normalized mutual information is defined as

$$NMI(U,V) = \frac{H(U) + H(V)}{H(U,V)} = \frac{MI(U,V)}{H(U,V)} + 1$$

Information theory offers a host of other measures that can potentially be used for registration. A few such measures are discussed in [38].

1.2.4 Iterative Optimization

The similarity measure employed during registration is generally optimized via an iterative strategy. Several such optimization strategies have been discussed in the literature and can be primarily classified into two broad categories – gradient-based and non-gradient-based [39]. Gradient-based strategies include techniques such as steepest gradient descent, conjugate-gradient methods, quasi-Newton methods and Levenberg-Marquardt optimization. The speed of convergence of such methods can be improved by using stochastic approximation and importance sampling techniques [40]. Downhill

simplex and Powell's direction set method are examples of non-gradient-based strategies. The registrations in this thesis use the downhill simplex optimization strategy proposed by Nelder and Mead [41]. The Nelder-Mead Simplex method is relatively robust to local minima and their effectiveness in multiresolution MI-based registration schemes have been shown in [39].

1.3 Summary of Contributions

Standard techniques for tumor response assessment, like RECIST 1.1 or WHO criteria, make many assumptions about the geometry and shape evolution of the tumor which limits their efficacy in measuring small early changes. Most current research efforts advocate the use of sequential 3D segmentations to measure tumor volume change. This thesis explores the use of image registration-based methods for tumor volume change measurement and investigates if such methods can yield accurate and low noise estimates of tumor volume change.

The following is a summary of the contributions of this thesis.

- 3D nonlinear image registration has been used to develop a low noise, low bias tumor volume change measurement algorithm.
- A linear mixed effects model has been used to characterize the performance of the proposed algorithm. It is shown that the error in measuring volume change increases with increase in the volume of the tumor and decreases with the increase in its normalized mutual information content even when that is not the similarity measure being optimized.

- The developed registration-based change measurement algorithm has been compared with other approaches and it has the potential to outperform indirect segmentation-based change measurement methods.
- The potential of an accurate registration-based volume change measurement algorithm in tracking progression of chronic obstructive pulmonary disease has been suggested in an initial study.

1.4 Dissertation Overview

This chapter introduced the problem, provided an overview of image registration and summarized the contributions of this thesis. Chapter 2 describes a preliminary study aimed at assessing the feasibility of image registration-based algorithms for measuring volume change in tumors. Two registration-based change measurement methods are discussed and tested on hepatic lesions in interval CT exams and some of the concepts developed here are investigated in later chapters. Chapter 3 discusses the proposed 3D warping registration-based tumor volume change measurement algorithm. The bias and variance of the method are estimated by testing in vivo and synthetic breast tumors in diffusion MRI exams where truth is unequivocally known. Statistical models are developed to characterize the performance accuracy of the proposed algorithm. Chapter 4 discusses the major segmentation-based approaches for measuring volume change in tumors and demonstrates some of the challenges faced by such approaches. The developed registration-based method for estimating tumor volume change is compared with two segmentation-based approaches on common datasets. Chapter 5 extends the application of the developed registration-based tumor volume change measurement

algorithm to investigate its potential promise in tracking chronic obstructive pulmonary disease. Chapter 6 discusses the contributions of this thesis and future work.

Chapter 2

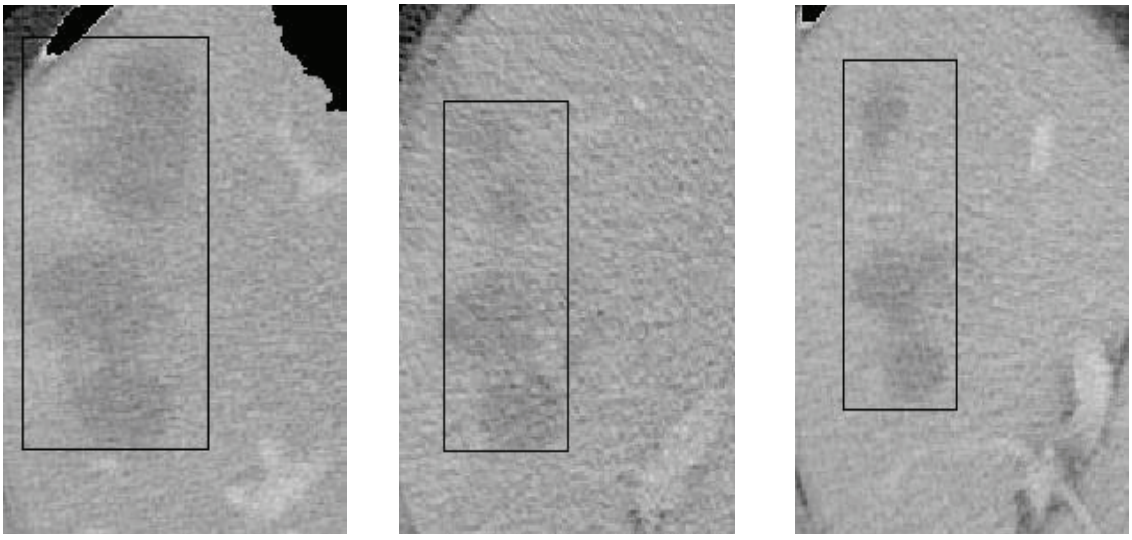
A preliminary study to assess the feasibility of image registration-based methods for lesion volume change estimation

2.1 Introduction

This chapter discusses a pilot study conducted on hepatic lesions in interval CT exams to assess the feasibility of image registration-based methods for lesion volume change estimation. Rather than a rigorous analysis of such methods, this was essentially a study of concepts. Some valuable ideas developed in this preliminary study are explored further in later chapters. This chapter is based on Sarkar et al. [42].

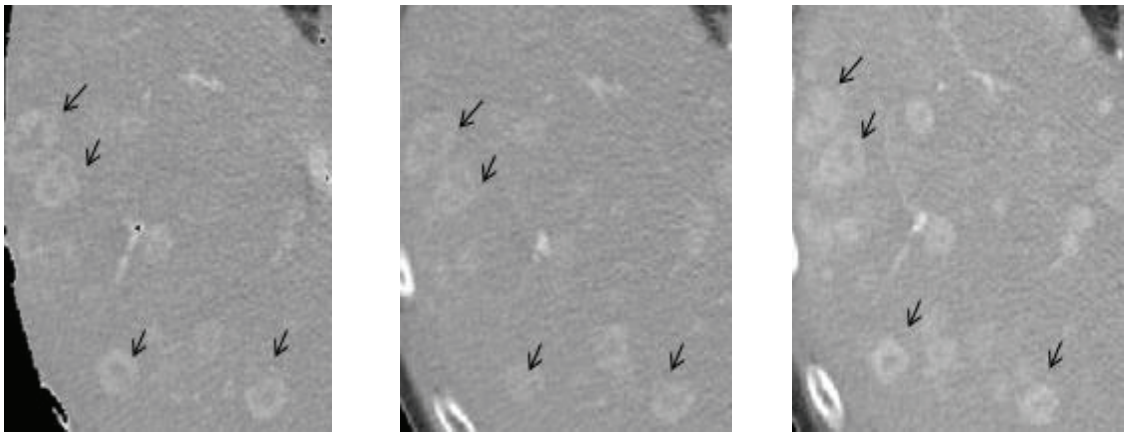
The liver is the largest gland of the human body and serves a wide range of vital functions including detoxification, protein synthesis and metabolism. It is also a common site for malignant tumors – both primary and metastatic. The National Cancer Institute reports an increase in incidence and mortality rates for liver cancer over the past two decades with an estimated 18,910 deaths in the U.S. in 2010 alone [43]. Typical treatment options include one or a combination of surgery, ablation, chemotherapy or radiation therapy. During the course of therapy, diagnostic CT scans of the patient's liver are acquired at regular intervals to monitor the response to therapy over time. Standard clinical radiological readings, for reasons discussed in Chapter 1, often prove inadequate

for response evaluation. An accurate quantitative measure of tumor volume change over interval scans, however, can considerably aid in ascertaining the efficacy of ongoing therapy. The need then is to develop techniques for accurately measuring tumor volume change between interval exams and this is currently considered a significant challenge. Figures 2-1 and 2-2 illustrate the complexity of the problem.



(a) (b) (c)

Figure 2-1 CT liver scans of patient showing tumor (a) baseline (b) after 1 month (c) after 5 months



(a) (b) (c)

Figure 2-2 CT liver scans of patient showing gross diffuse metastatic disease (a) baseline (b) after 3 months (c) after 8 months

Shown above are interval CT liver scans of two patients, chronologically arranged from left to right. The tumors have been marked for the benefit of the reader. The tumor depicted in Figure 2-1 has complex fuzzy boundaries and the complicated shape evolution over time is also evident from the scans. Figure 2-2 shows a patient with gross diffuse metastatic disease. The tumors have diffused throughout the organ and have dark necrotic cores surrounded by active tissue evident from the wide intensity variations within the tumors. As can be ascertained from the figures, quantification of volume change across intervals for in vivo lesions can be a complex task.

An IRBMED (Institutional Review Board of the University of Michigan Medical School) approved pilot study was conducted on human subjects to assess the feasibility of using registration-based approaches for measuring volume change of hepatic tumors in interval CT exams. The de-identified retrospective data that was analyzed was obtained from interval abdominal CT exams of high risk patients diagnosed with stable disease or

partial response to therapy. Two registration-based approaches – Weighted Gray Value Difference and Non-linear warping followed by Jacobian integration – were employed to measure the volume change of lesions across interval exams and the results were compared to estimates obtained through careful manual segmentation of the lesions in the exams.

The use of image registration-based techniques for tumor volume change estimation is relatively less explored as compared to sequential segmentation [19, 20]. An initial registration-based endeavor to quantify tumor volume change was made by Thirion et al. [44]. Thirion registered several synthetic multiple sclerosis lesions using the Demons nonrigid registration algorithm [45] and calculated tumor volume change by integrating the deformation field using concentric spheres and embedded isosurfaces. Reeves et al. [46] used 3D rigid body registration to improve the consistency of sequential segmentations and the accuracy of change measurements in interval scans of pulmonary nodules. Meyer et al. [47] used low degrees of freedom registration to align liver lesions. This was followed by a weighted gray value subtraction scheme to compute lesion volume change. Extensions to the Meyer method and modifications to Thirion's scheme including the use of a spatially varying Jacobian map of local scale changes will be discussed in this chapter.

2.2 Registration-based methods for Volume Change Measurement

A total of six lesion pairs in CT liver scans of high-risk cancer patients with scan intervals ranging from 1 month to 4 years were examined. The 3D experimental 512 x 512 de-identified CT scans were acquired using a GE Medical, LightSpeed Ultra CT

scanner. The voxel volumes ranged from 2.2 to 3.7 mm^3 with slice thicknesses of 5 mm and peak kilovoltages of 120 kVp. To obtain an estimate of the "true" lesion volume difference between the interval scans, all the lesions were carefully segmented in every exam by S. Sarkar. These manual segmentations were repeated 10 times independently to obtain the lesion volume difference between each homologous and reference exam pair. The lesions were then analyzed using two registration-based methods, Non-linear warping followed by Jacobian integration and Weighted Gray Value Difference. All techniques were implemented using MIAMI-Fuse (Mutual Information for Automatic Multimodality Image Fusion), a software program for the automatic alignment of medical datasets obtained using same or different modalities [36]. Both registration-based techniques were used to obtain 10 independent estimates of the volume change in each lesion pair. Sample means, sample standard deviations and 95% confidence intervals were estimated using a bias-corrected and accelerated Bootstrap method [48].

Let us now look at the two registration-based methods in detail.

2.2.1 Weighted Gray Value Difference

This method uses a weighted gray value difference image of two registered interval exams to estimate the change in lesion volume of homogeneous lesions only. It involves a locally rigid registration between the homologous and the baseline or reference exam. Registration is followed by gray value subtraction of the registered scans yielding a difference image (Figure 2-3) of the lesion with improved contrast-to-noise ratio.

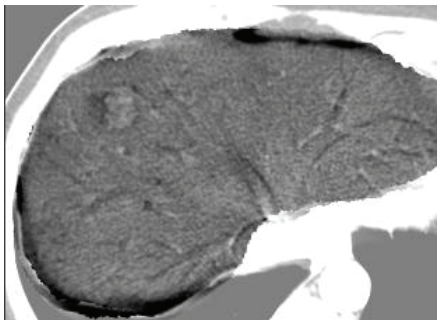


Figure 2-3 Difference image of registered homologous and reference scans

The change in volume is estimated by linearly weighting and trimming the histogram values obtained after manually segmenting the lesion in the difference image and multiplying the number of voxels obtained (full and partial) by the voxel volume. The linear weighting and trimming curve (shown in Figure 2-4) is essential to accurately account for the contribution of partial voxels as described in [47]. It should be noted that the histogram shown in Figure 2-4 is for an ideal case with no noise. A real histogram, affected by noise of the difference image, looks somewhat like the one shown in Figure 2-5.

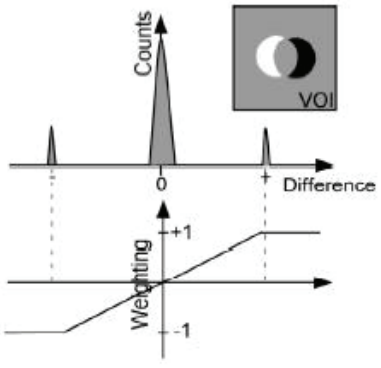


Figure 2-4 Example ideal histogram and trimming curve of a slightly misregistered lesion with no volume change

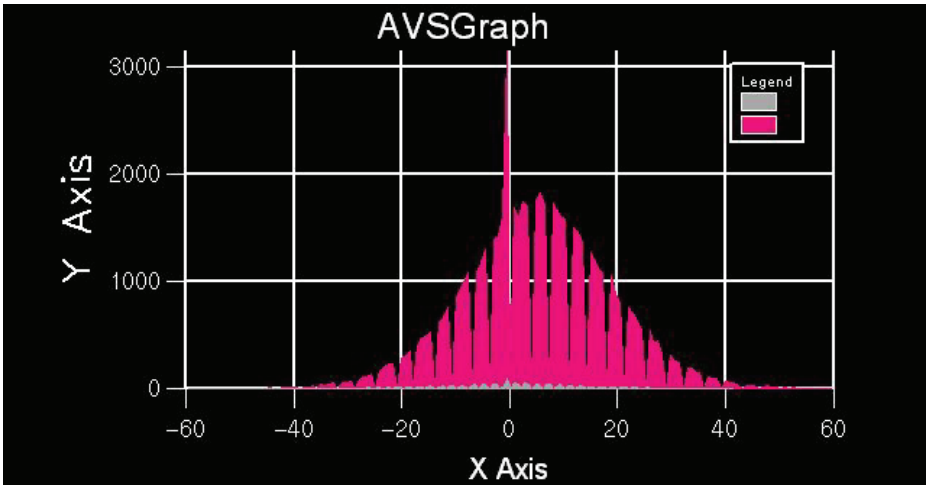


Figure 2-5 Histogram of lesion volume difference affected by noise

Meyer et al. [47] require the sketching of an additional VOI in the background, exclusive of the lesion, to combat contrast enhanced intensity differences in the subtracted image and set the zero difference offset for the weighting curve. This works fine if, between the reference and registered homologous exams, the difference in intensities of the two backgrounds is about the same as the difference in intensities of the two lesions. If this is not the case, then one would end up selecting an erroneous zero difference offset by drawing the additional VOI in the background. To counter this problem an approximate estimate of the lesion and background intensities are made in both exams prior to creating the difference image. If the difference is substantial, one of the exams is linearly scaled such that the mean background and lesion intensities of both exams overlap. A typical linear graph to do such a scaling is shown in Figure 2-6. The VOI to set the zero difference offset of the weighting curve is now defined on the difference image obtained through this modified registered reference-homologous pair.

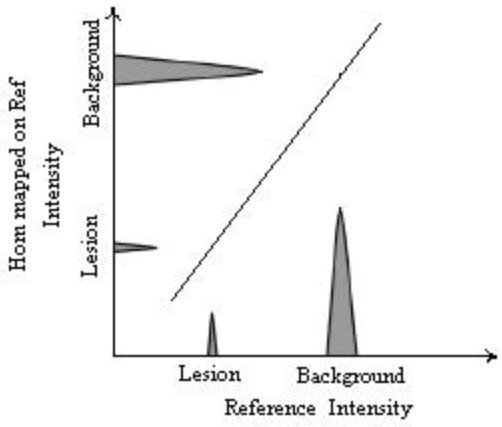


Figure 2-6 Linear scaling to match the background and lesion intensities in the exams before gray value subtraction

2.2.2 Non-linear warping followed by Jacobian integration

It is well known from the calculus of continuous variables e.g. tensors that the Jacobian determinant of a geometric transformation at a given voxel provides the local volume ratio between the deformed and the undeformed voxels. In this method, a non-linear registration scheme is employed to capture the deformation of the lesion over time and acquire a Jacobian map of tissue volume changes. The local scale changes obtained from the map are integrated to yield the net volume change of the lesion. If the reference is feature-poor and thus low on mutual information outside the lesion, as is often the case in hepatic CT images, constraints are imposed on the Jacobian to ensure local volume preservation beyond the tumor boundaries. Such a method is potentially capable of tracking complicated shape changes in the lesion and providing accurate volume change estimates.

Consider a reference scan that is feature-poor outside the lesion (Figure 2-7). Since there are no intrinsic features to drive the registration outside the lesion in the feature poor surround, constraints need to be imposed on the deformation interpolant to ensure local volume preservation outside the lesion while not affecting the volume change calculations within the lesion. This is ensured by the following implementation. Initially, a uniformly spaced grid of control points is positioned in the reference exam in 3D just outside the boundary of the lesion of interest. Figure 2-7 (blue points) shows the location of some of these points in one slice of the reference scan. Three control points are placed in the homologous exam approximately corresponding to the locations of the first three distance sorted control points in the reference exam. A mutual information guided rotate-translate registration scheme is then employed to obtain the transformation between the exams. This transformation is then used to place the corresponding grid of points in the homologous exam. Next, another uniformly spaced grid of points is placed on the reference, this time lining or just within the lesion boundary as shown in Figure 2-7 (red points). The control points on the lesion in the reference exam are distance sorted and six points are placed on the lesion in the homologous exam that correspond to the approximate locations of the first six points on the reference lesion. Note that a uniform grid of points, as deployed here, is not generally an optimal distribution of the control points. The density and distribution of control points should be based on what the local mutual information content of the object of interest allows it to support [49].

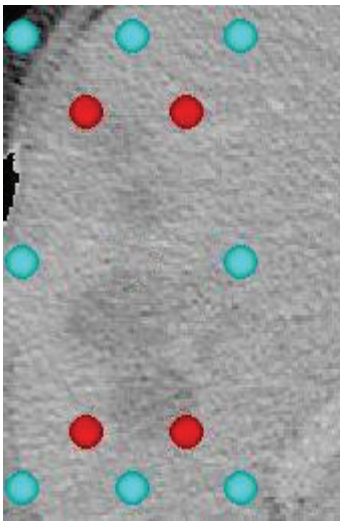


Figure 2-7 Placement of outer and inner grid of control points on reference image

The reference and homologous exams are now automatically non-linearly registered using thin plate splines (TPS) as the deformation interpolant with the chosen six points acting as the starting point. The objective function used for the registration is mutual information. During registration it is ensured that the homologous outer grid of points *do not move* from their fixed locations obtained through the rotate translate registration in the previous step. This means that, in the registration process, the optimizer (Nelder-Mead Simplex) *only* controls the inner grid of homologous points to compute the deformation. The resulting vector-valued deformation field F is used to obtain a Jacobian map representing the local volume expansion or contraction of every voxel (x, y, z) computed by the warp. The Jacobian determinant at a given voxel (x, y, z) can be computed as shown below.

$$J_F(x,y,z) = \begin{vmatrix} \frac{\partial F_x(x,y,z)}{\partial x} & \frac{\partial F_x(x,y,z)}{\partial y} & \frac{\partial F_x(x,y,z)}{\partial z} \\ \frac{\partial F_y(x,y,z)}{\partial x} & \frac{\partial F_y(x,y,z)}{\partial y} & \frac{\partial F_y(x,y,z)}{\partial z} \\ \frac{\partial F_z(x,y,z)}{\partial x} & \frac{\partial F_z(x,y,z)}{\partial y} & \frac{\partial F_z(x,y,z)}{\partial z} \end{vmatrix}$$

Local expansions and contractions yield positive Jacobian values $J_F(x, y, z) > 1$ and < 1 respectively, whereas a value of 1 indicates no volume change in the voxel. The volume change is estimated by summing up the local scale changes over the lesion as given by the Jacobian. The Jacobian values are monitored throughout the registration to prevent any folding and hence obtain a bijective mapping between the exams. The crucial step of tethering the outer grid of points during non-linear registration constrains the effect of the thin plate splines and ensures that the Jacobian values go back to 1 outside the lesion (ensuring local volume preservation outside the tumor). The use of TPS ensures this transition is maximally smooth [27]. The vector valued TPS displacement in 3D, $F(x)$, is given by the equation below.

$$F(x) = Ax + B + I \sum_{i=1}^{N} W_i r_i$$

Here A and B define an affine transformation and I is the identity matrix. A, B and the TPS coefficients $\{W_i\}$ are determined by solving the above set of linear equations at the N control point locations in the reference as discussed in Chapter 1. The TPS is a globally supported function and hence computationally complex. Also, large deformations can lead to singularities in the matrix leading to folding. This is prevented, in the experiments, by removing control points (on the inner grid only) from the locations nearest the most negative Jacobian iteratively during the registration.

2.3 Results

The mean Homologous - Reference lesion volume differences (in cc) together with 95% confidence intervals are listed below for the six lesion pairs for all three techniques –

manual segmentation (MS), Weighted Gray Value Difference (WGVD) and Non-linear warping followed by Jacobian integration (JI) (Table 2-1).

	MS	**JI**	**WGVD**
Lesion 1 (Reference Tumor Volume: 2.5 cc)			
Mean Vol. Change	6.724	7.149	6.651
Mean CI (95%)	(6.702, 6.751)	(6.718, 7.400)	(6.549, 6.728)
Lesion 2 (Reference Tumor Volume: 153.9 cc)			
Mean Vol. Change	-67.124	-74.184	-43.816
Mean CI (95%)	(-64.909, 69.524)	(-67.155, -80.401)	(-37.690, -50.506)
Lesion 3 (Reference Tumor Volume: 153.9 cc)			
Mean Vol. Change	-41.449	-34.534	-13.833
Mean CI (95%)	(-40.317, -42.809)	(-32.052, -36.435)	(-12.375, -15.640)
Lesion 4 (Reference Tumor Volume: 5.1 cc)			
Mean Vol. Change	0.950	1.037	0.864
Mean CI (95%)	(0.890, 0.988)	(0.738, 1.281)	(0.847, 0.872)
Lesion 5 (Reference Tumor Volume: 5.1 cc)			
Mean Vol. Change	-0.024	0.086	-0.125
Mean CI (95%)	(-0.109, 0.062)	(-0.141, 0.272)	(-0.235, 0.019)
Lesion 6 (Reference Tumor Volume: 12.3 cc)			
Mean Vol. Change	-5.701	-8.350	-4.137
Mean CI (95%)	(-5.442, -5.960)	(-7.430, -9.240)	(-4.069, -4.174)

Table 2-1 Mean volume change in cc together with 95% confidence intervals for all 6 lesions using all three methods

Figure 2-8 and Figure 2-9 are plots of mean lesion volume change estimates together with 95% confidence intervals for the two algorithms in comparison to that of manual segmentation. In accordance with the observations of Meyer et al. [50], the higher variance or wider 95% confidence intervals associated with greater tumor size can be noted for the manual segmentations shown in the plots (reference tumor volumes are included in Table 2-1). In both figures small boxes inside the main plot have been provided to offer a magnified view of the relatively smaller changes. It should be noted that ground truth volume change is not known in this study. This necessitates the reliance on careful manual segmentations for obtaining estimates of the "true" or "close to true" volume change. The dashed line across the diagonal can be viewed as an indicator of agreement between the volume change estimated by the two registration-based methods and the "true" volume change obtained via manual segmentation. For good agreement, the estimated 95% confidence interval of volume change using the algorithms should intersect this dashed line. In other words, good agreement would imply there is an overlap in the 95% confidence intervals of volume change as estimated by manual segmentation and as estimated by the algorithms. From the plots this overlap is evident in 2 out of 6 cases for the weighted gray value difference method and in 4 out of 6 cases for the non-linear warping followed by Jacobian integration method (see Figure 2-8 and Figure 2-9 and Table 2-1).

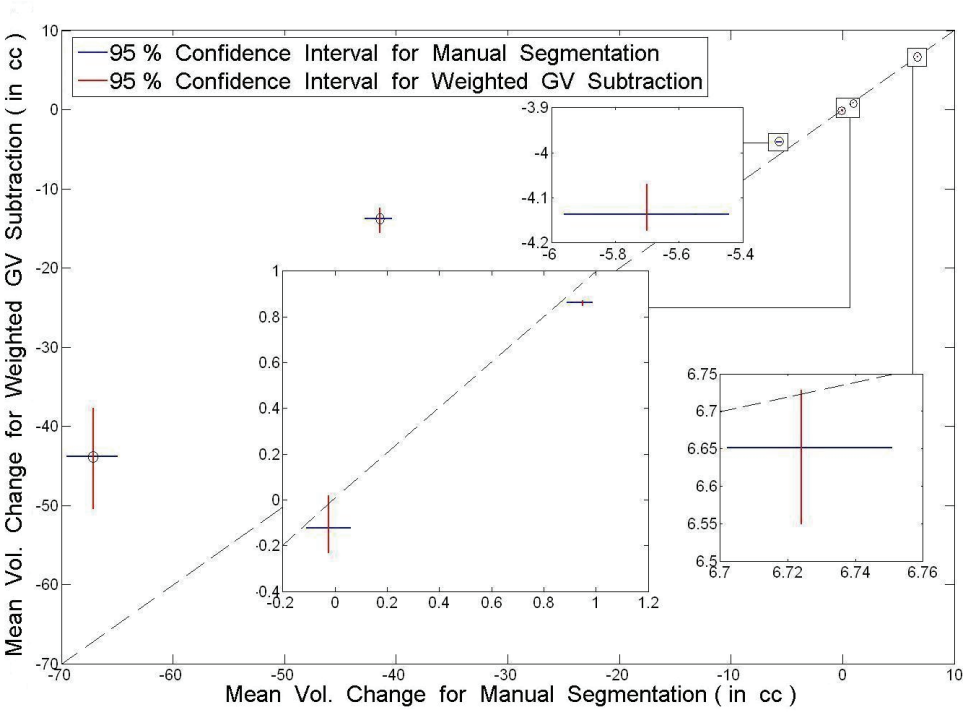


Figure 2-8 Mean Vol. Change for Weighted GV Subtraction vs Manual Segmentation. From left to right: Lesion number 2, 3, 6, 5, 4, 1. Reference Tumor volumes are included in Table 2-1

The overall accuracy of the non-linear warping method is clearly better compared to the weighted gray value difference method especially for large effect sizes (large volume changes) as evidenced by Figure 2-8, Figure 2-9 and Figure 2-10. This is likely due to the inhomogeneities typically associated with large volume changes in large lesions where the simple linear weighting scheme of the weighted difference method proves ineffective.

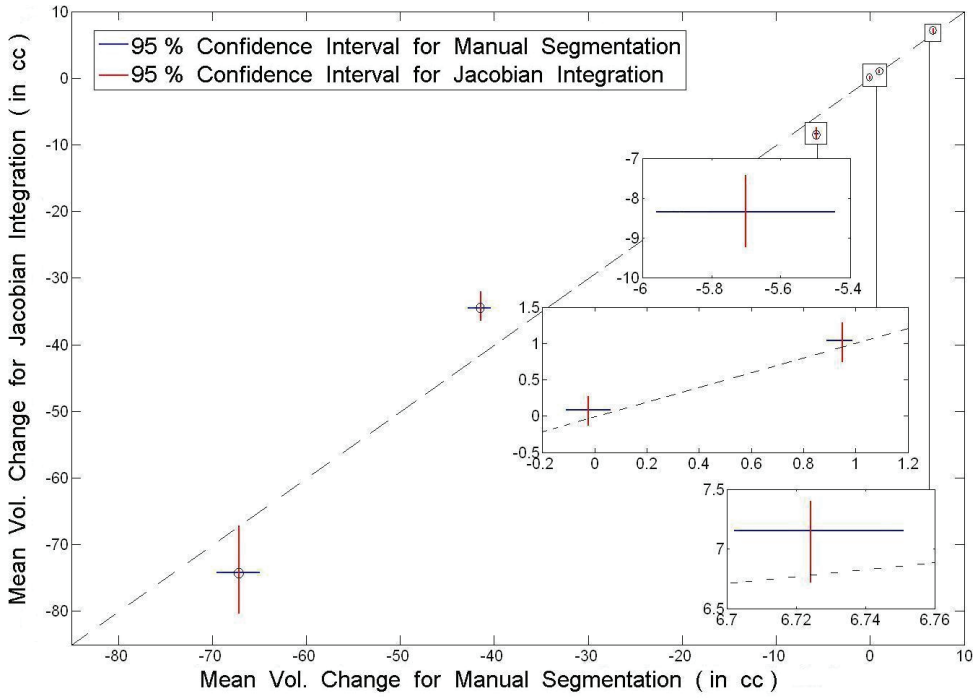


Figure 2-9 Mean Vol. Change for Jacobian Integration vs Manual Segmentation. From left to right: Lesion number 2, 3, 6, 5, 4, 1. Reference Tumor volumes are included in Table 2-1

Figure 2-11, Figure 2-12 and Figure 2-13 show checkerboard plots of non-linear warping registration for three lesions. The region of focus of the registration is outlined in the figures and the good overall alignment in these regions can be noted from the figures.

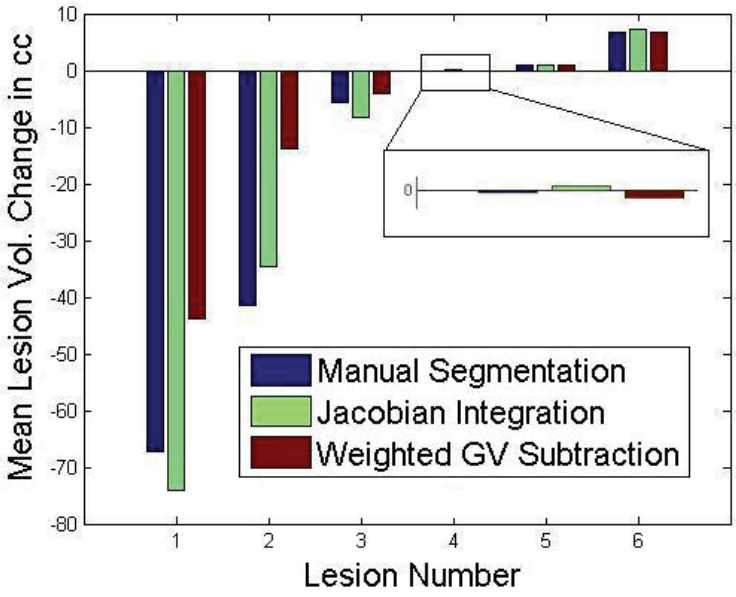


Figure 2-10 Mean lesion volume change in cc as estimated by manual segmentation, non-linear warping followed by Jacobian integration and weighted gray value difference

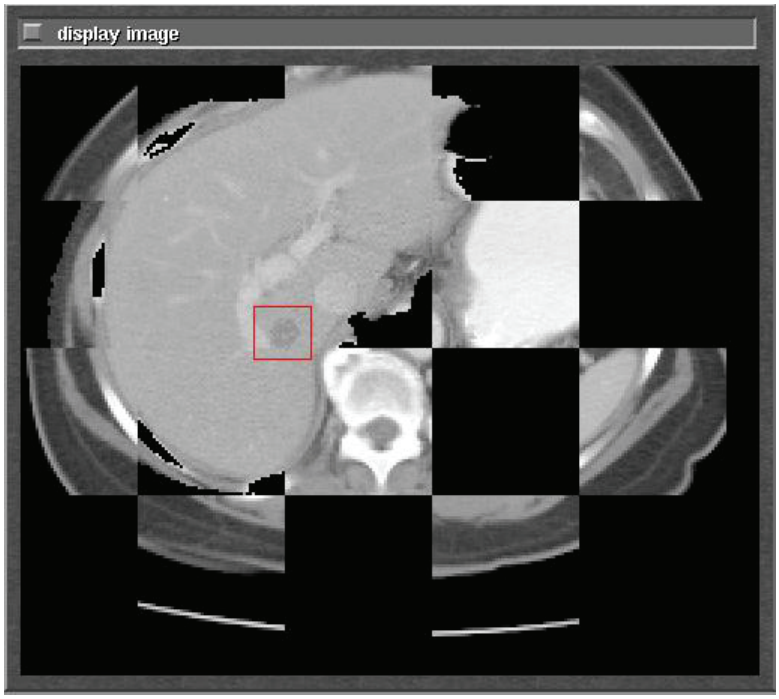


Figure 2-11 Checkerboard plot of non-linear warping registration for Lesion 1

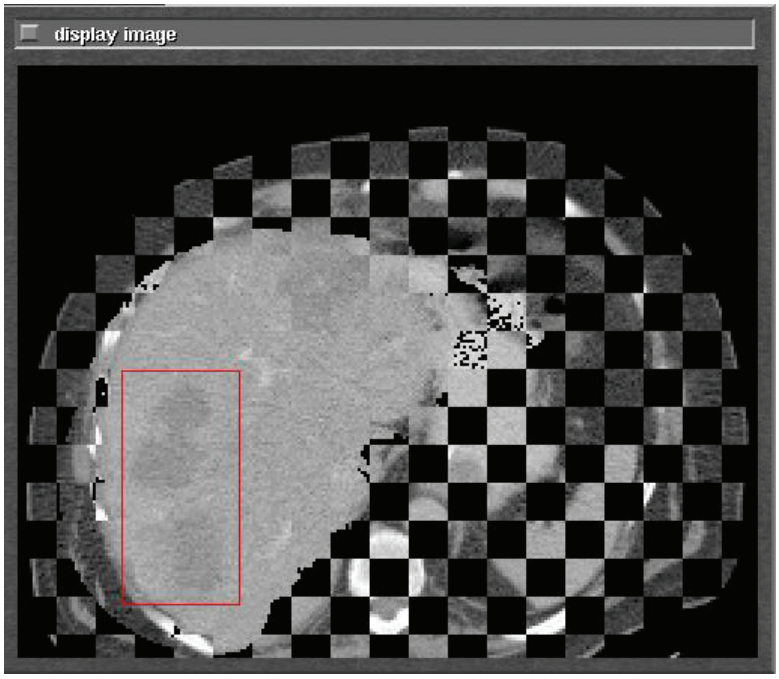


Figure 2-12 Checkerboard plot of non-linear warping registration for Lesion 2

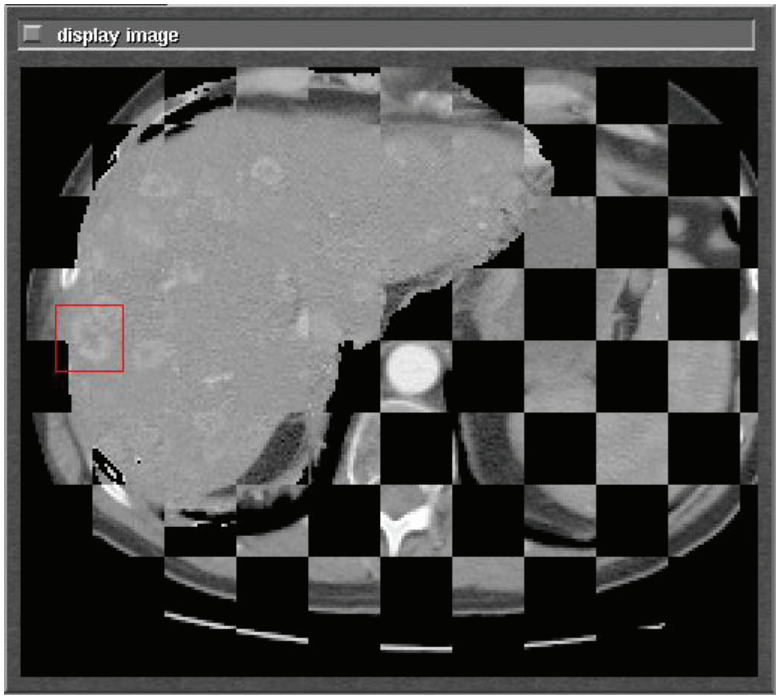


Figure 2-13 Checkerboard plot of non-linear warping registration for Lesion 4

2.4 Discussion

This pilot study demonstrated the potential promise of registration-based algorithms in measuring tumor volume change. The Weighted Gray Value Difference method is by design insensitive to slight misregistration but it is, however, sensitive to the lesion model used. Presently it simplistically assumes that the lesion of interest is a simple lesion with uniform contrast. In the case of inhomogeneous lesions the resulting difference histogram ends up with multiple peaks and the simple linear weighting scheme breaks down. While it was accurate in the case of small, simple, homogeneous lesions it is unlikely to fare well when these assumptions are violated. Non-linear warping followed by Jacobian integration, on the other hand, is potentially capable of tracking complicated shape changes as well as intensity variations in the lesion and does not make assumptions about the lesion model. Based on the results of this initial study, future efforts were directed towards improving and assessing the accuracy of measuring tumor volume change using non-linear registration only.

Another interesting concept developed in this chapter is the outer 'fence' of control points that is deployed in the feature poor reference dataset to enable the Jacobian to go back to 1 immediately beyond the lesion. This is a useful trick for handling datasets that have low mutual information content beyond the tumor boundaries. Since subsequent investigations on different datasets did not encounter this problem associated with low mutual information, this concept was not developed further during the course of this thesis. However, this is still potentially a useful idea and more work needs to be done to ascertain the optimal density and location of such a grid of points with respect to the inner grid of points placed on the tumor.

Chapter 3

Evaluation of an image registration-based algorithm for measuring tumor volume change

3.1 Introduction

In this chapter a 3D warping registration-based algorithm is developed to estimate volume change in tumors. The tumor of interest is spatially registered between two interval scans and the volumetric deviation of the tumor is calculated by summing local changes in volume obtained from a spatially varying Jacobian map. Such an approach not only provides a summary statistic of volume change but can also potentially show regions of differential growth and contraction across the lesion. Also the incorporation of partial volume effects, often ignored in some segmentation approaches, is automatically included in a registration-based paradigm.

The algorithm best suited for early volume change detection would be the one with the least measurement noise and least bias on average [21]. One major obstacle in the assessment or validation of different volume change measurement algorithms is the lack of sources where true volume change is reliably known. Synthetic exams and very short interval exams provide solutions to this problem. The bias and noise variance of the proposed registration algorithm, in the quantification of volume change in breast tumors, is assessed using such sources.

3.2 Materials and Methods

This section describes the dataset and the registration algorithm used to estimate tumor volume change together with the statistical models used to characterize the performance of the algorithm.

3.2.1 Dataset

The clinical merit of a volume change measurement algorithm will be characterized by its ability to evaluate response to therapy *early*, i.e., by its ability to measure small volume changes that occur in short time intervals. In other words from the clinical standpoint the main motivation lies in evaluating the noise and bias of an algorithm in measuring zero or close to zero tumor volume change. Although large non-zero volume changes may be easily discerned by an expert radiologist, a good volume change measurement algorithm should also be able to track small changes. To gauge the performance of the registration-based change measurement algorithm that will be discussed below, two kinds of interval scans were used – clinical interval scans where true tumor volume change is known to be zero for assessing the repeatability of the method and synthetic interval scans with random but known non-zero tumor volume change primarily for observing functional bias effects.

The clinical dataset used for this investigation comes from an ongoing study at the University of Michigan aimed at early evaluation of response to neoadjuvant chemotherapy in 9 breast cancer patients. In the imaging protocol every patient receives two pre-treatment baseline isotropic diffusion MRI (dMRI) scans within a very short interval (15 minutes). These exams are also referred to as “coffee-break” exams. An

expert radiologist delineates the tumor's volume of interest (VOI) using a contrast enhanced anatomical scan. The enhanced anatomical scan is warped onto one exam out of each "coffee-break" pair and the VOI mask is subsequently mapped onto the associated diffusion scan. Since these diffusion exams are taken within a few minutes of each other, it is reasonable to assume that within this interval there are no macroscopic changes to the tumor, i.e., the ground truth of tumor volume change between these two exams is zero. However, even though both scans were derived from the same volume, they differ due to the decorrelating effects of noise during scan acquisition, patient repositioning and patient motion. These factors make these very short-interval exams excellent resources for assessing the measurement noise associated with volume change measurement algorithms at the null point of volume change. Manual segmentations, even when done by experts, have significant variability [50]. It is therefore likely that the tumor volumes as represented by the expert in the nine patients are not entirely accurate. However, this does not affect the way the algorithm's performance is evaluated because the study's assessment paradigm considers changes in tumor volume and not the individual tumor volumes as ground truth. Each "coffee-break" pair represents the same volume in the patient – any non-zero deviation measured by the algorithm is in error. Another point to note is that volume change is not computed here via subtraction of multiple expert annotated volumes which would add to the variability of change measurements. Only one annotation is done, that on the reference volume and the volume change is computed by summing the Jacobian of the deformation over this expert annotated reference volume of support. All "coffee-break" MR datasets used in the study are 288 x 288 x 60 with a voxel volume of 3.84 mm^3 (1.25 mm x 1.25 mm x 2.46 mm);

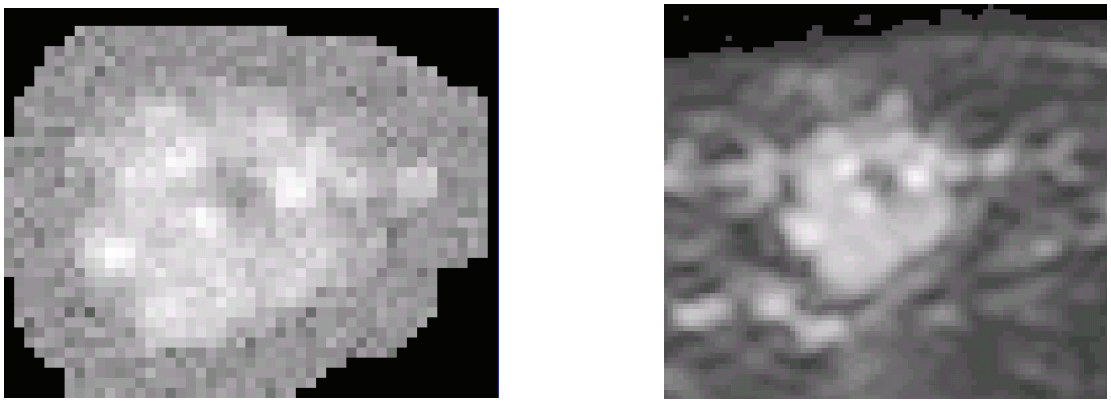
the registrations are performed on the fat suppressed, zero diffusion gradient image, i.e., the T2-weighted, b=0 image, acquired using a 7-coil breast receiver array on a Philips 3T magnet, Tr = 4720 ms, Te=50 ms, sense=2. The tumor volumes across the nine patients ranged from 5.39 cc to 57.35 cc as defined by the expert.

While the "coffee-break" dataset provides realistic noise at the operating point of zero tumor volume change, it is difficult to obtain clinical scans with non-zero volume change where truth is also unequivocally known. To counter this problem, data sources with known non-zero tumor volume change can be obtained from synthetic scans like phantoms [51] or mathematical models [52]. In this study we synthesized 30 homologous scans by 3D random scaling, rotating, translating and linearly interpolating one scan of a breast cancer patient (see equation below) chosen from the pool described above.

$$H = R_x(\varphi)R_y(\alpha)R_z(\theta)sC + T(t_x, t_y, t_z)$$

In the above equation H is the coordinate vector of the generated synthetic scan, C is the coordinate vector of the original clinical scan, s is the isotropic scaling factor chosen randomly from a uniform distribution to create contractions and expansions of the original volume in the range (-80%, 80%), θ, α and φ are yaw, pitch and roll angles respectively to generate 3D rotations and are chosen randomly from $\mathcal{N}(0\ degree, 25\ square\ degrees)$, t_x, t_y and t_z are translations in x, y and z directions respectively and chosen randomly from $\mathcal{N}(0\ mm, 4\ mm^2)$. The random rotations and translations caused the tumor in some synthetic scans to end up outside the field of view. Such scans were discarded. The radiologist's segmented tumor mask (VOI) in the original scan was dilated by 3 voxels such that the tumor edge voxels play a role in the

subsequent registration. In decorrelating the noise in the synthetic exams from the reference region of interest, 30 reference scans were derived from the masked scan by adding a random Gaussian noise realization which led to a five-fold decrease in SNR. Each pair in the resulting set of 30 homologous and reference scan pairs now had a random, known, non-zero volume change in the region of interest. One such pair is shown in Figure 3-1.



(a) (b)

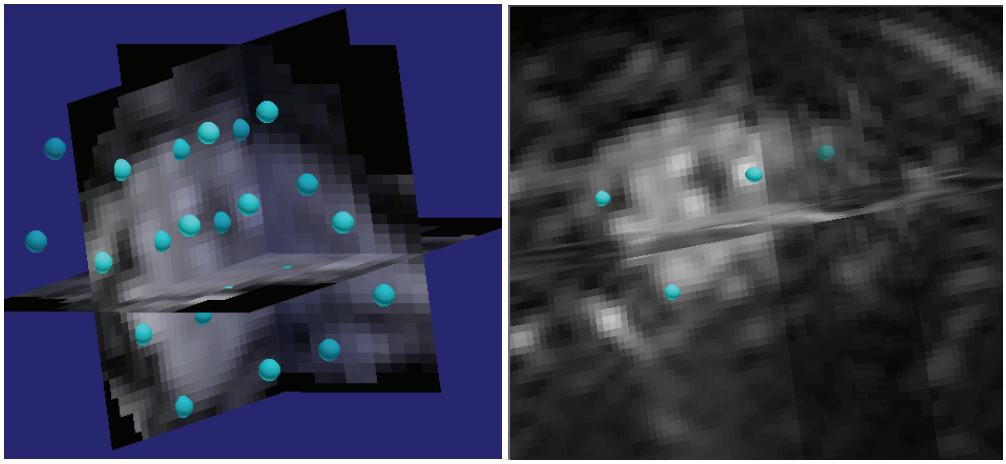
Figure 3-1 Sample slice from a synthetically generated dataset (a) slice of dilated reference VOI with added, decorrelated noise; masked volume of interest was 42.19 cc (b) scaled, rotated, translated and interpolated homologous image

3.2.2 Registration Algorithm

The tumor of interest is registered between two interval scans using a 3D warping registration algorithm. Warping is necessary to accurately capture the potentially heterogeneous deformation of tumor compartments in the breast. All registrations were implemented using MIAMI-Fuse (Mutual Information for Automatic Multimodality Image Fusion), a software program for the automatic alignment of medical datasets obtained using same or different modalities [36].

The registration details for the "coffee-break" exams, i.e., exams where true tumor volume change is known to be zero, is described below. For each of the 9 patients:

- The tumor's volume of interest as drawn by the radiologist was dilated using a 3x3 binary structuring element to include regions external to the tumor. The dilated region served as the reference volume and ensured that edge information associated with the tumor was not lost while still keeping the focus of registration on the tumor. Three different sets of dilations (approximately 50%, 100% and 200% of the size of the segmented tumor volume) were chosen to investigate if incorporating information external to the tumor improved the registration of the tumor itself.
- A set of hexagonally close-packed control points were automatically generated in the 3D reference volume spaced 12 mm apart (Figure 3-2 (a)). Small variations to this spacing were not found to affect registration accuracy and hence this spacing was used for all the tumors. However, in general, the number N (e.g. in 3D, degrees of freedom = 3N) or spacing of control points chosen should be based on what degrees of freedom the mutual information density of the region of interest supports in practice without folding. The control points were then distance-sorted in the reference volume by maximizing the Euclidean distance between the points. This ordered the control points in a coarse to fine scale across the reference volume to set up a decreasing scale space registration paradigm.



(a) (b)

Figure 3-2 Placement of control points (a) automatic hexagonal arrangement of control points in reference volume (b) manual placement of first 5 corresponding control points (only 4 are visible here) in the homologous volume

- For each of the first five distance-sorted control points in the reference, one approximate corresponding control point was chosen in the homologous volume (the other "coffee-break" scan) as shown in Figure 3-2 (b). Except for this manual initialization all the other steps described above and below were automatically controlled by MIAMI Fuse. Five control points were required for starting because less than five describes only a linear solution which is unlikely to lead to capture of a warping required to register the deformed tumor after repositioning in the "coffee-break" exam.
- A 3D warping registration was performed between the two volumes to align the homologous image with the reference image. The algorithm implemented an iteratively increasing degrees of freedom and decreasing scale space warping registration. Two similarity measures were separately used for registration - mutual information and normalized cross correlation. Mutual information is robust to

contrast enhanced intensity differences as opposed to some voxel similarity measures such as sum of squared intensity differences, sum of absolute differences and normalized cross correlation. So similar structures with different contrasts just show up as different clusters in the bivariate histogram and can still be registered using mutual information. Normalized cross correlation has been used for intra-modality registration and is optimum when the much stricter assumption of a linear relationship between the intensity values of the two images is valid [23]. The deformation interpolant used for warping was thin plate splines (TPS). In the absence of any knowledge regarding the mechanical properties of the tissue, TPS was used as the deformation interpolant as it is maximally smooth [27]. To prevent the deformation from folding, the sign of the Jacobian determinant was monitored at the finish of each schedule line optimization. The control point closest to the most negative Jacobian (if there were any) was iteratively removed after each optimization cycle and the cycle repeated until there were no negative Jacobian values. The optimizer used was Nelder-Mead Simplex which is relatively robust to local minima [53].

- The determinant of the first order partial derivatives of the deformation tensor F obtained by the warping registration was used to form a Jacobian map representing the local volume expansion or contraction of every voxel (x, y, z) in the reference volume. The Jacobian determinant at a given voxel (x, y, z) can be computed as shown below.

$$J_F(x,y,z) = \begin{vmatrix} \frac{\partial F_x(x,y,z)}{\partial x} & \frac{\partial F_x(x,y,z)}{\partial y} & \frac{\partial F_x(x,y,z)}{\partial z} \\ \frac{\partial F_y(x,y,z)}{\partial x} & \frac{\partial F_y(x,y,z)}{\partial y} & \frac{\partial F_y(x,y,z)}{\partial z} \\ \frac{\partial F_z(x,y,z)}{\partial x} & \frac{\partial F_z(x,y,z)}{\partial y} & \frac{\partial F_z(x,y,z)}{\partial z} \end{vmatrix}$$

Local expansions and contractions yield positive Jacobian values $J_F(x,y,z) > 1$ and < 1 respectively in a Jacobian map, whereas a value of 1 indicates no volume change in the voxel. The total volume change was estimated by summing up the local scale changes, as given by the Jacobian, over the volume annotated by the radiologist in the reference image as shown in the equation below.

$$\sum_{(x,y,z)\in U_{tumor}} (J_F(x,y,z) - 1)v$$

Here U_{tumor} represents the 3D tumor volume as annotated by the radiologist in the reference scan and v represents the voxel volume of the reference exam.

Now the registration details for the synthetic exams with known non-zero volume change will be described. The registrations of the 30 synthetic pairs of homologous and reference scans followed the same process described above except only one similarity metric (mutual information) and one dilation (200% tumor volume) was used. All 30 TPS warping registrations were automatically started with the same 64 distance-sorted hexagonally close-packed control points placed on the reference scan and the corresponding guess for the first 5 out of the 64 control points placed on the homologous scan. Note that due to the way these synthetic datasets were created, a rotate-isotropic scale-translate registration algorithm will likely provide a less noisy solution than a

higher degrees of freedom TPS fit. However the standard blinded approach was chosen to obtain a fair evaluation of the registration-based volume change measurement algorithm.

3.2.3 Statistical Model

Several linear mixed-effects (lme) models were fitted to the volume change results obtained from the "coffee-break" data. Selection of the model was based on the Bayesian information criteria [54]. Results obtained from using mutual information and that from using normalized cross correlation were fitted separately. The mixed effects models were implemented using the method of restricted maximum likelihood estimation via the lme function in the R package nlme and R's general purpose optimization routine. The following is the linear mixed effects model fitted to the data where the log transformed volume change error was considered as the response:

$$log\left(\left|V_{h_{ijk}} - V_{r_i}\right|\right) = log\left(V_{r_i}\right) + S_{ijk} + I_{ij} + \varepsilon_{ijk}$$

Interactions between predictors were not found to be significant. Another linear mixed effects model, shown below, was fitted to the data with the percentage volume change error as the response.

$$\frac{\left|V_{h_{ijk}} - V_{r_i}\right|}{V_{r_i}} * 100 = \theta + S_{ijk} + R_i + \varepsilon_{ijk}$$

Here $i = 1, 2, \ldots 9$ represents the nine tumors, j represents one of the 3 different dilations chosen and $k = 1,2, \ldots 10$ represents the different randomly restarted registration repetitions taken under each setting. V_h and V_r represent the estimated homologous tumor volumes and the manually segmented reference tumor volumes respectively. I is a

random tumor intercept for a given dilation nested within a given nodule, R is a random tumor intercept for a given nodule, θ is a fixed tumor intercept and ε is the random error term. S is the normalized mutual information content of the tumor – a similarity measure. Normalized mutual information was computed on the registered annotated tumor volume and is defined as the ratio of the sum of the entropies and joint entropy of the registered exams as shown below.

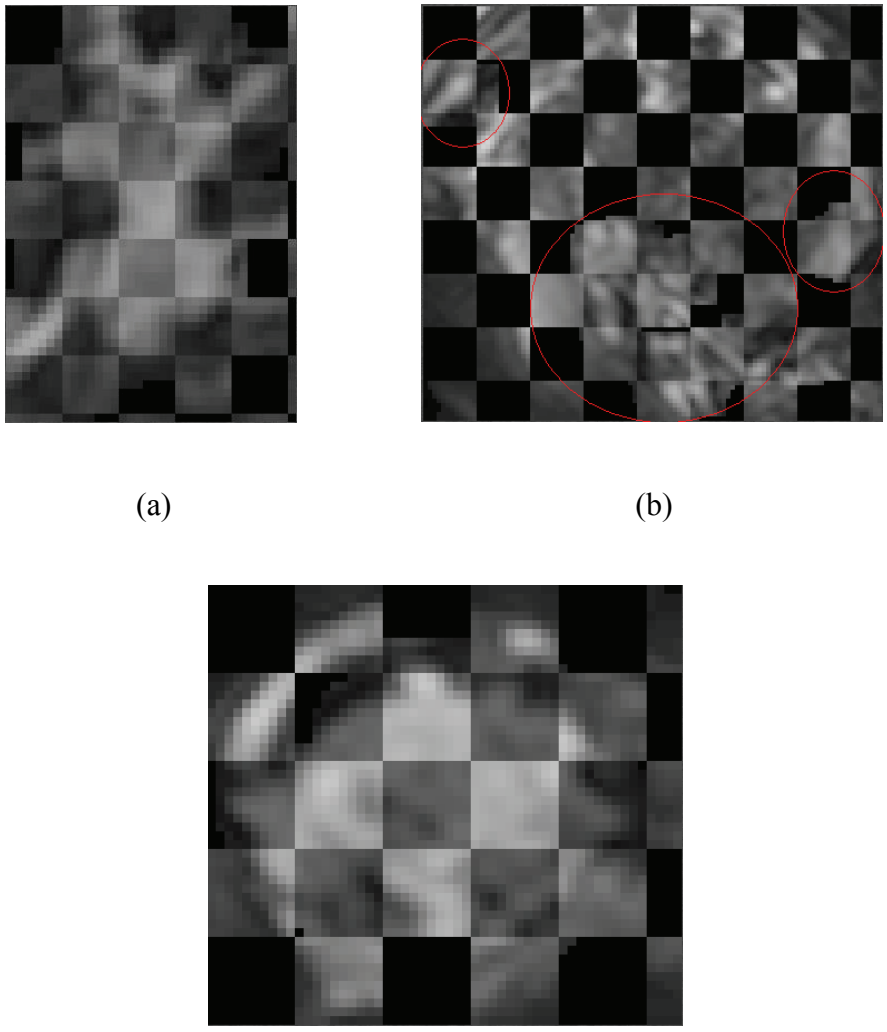
$$S(U,V) = \frac{H(U) + H(V)}{H(U,V)} = \frac{MI(U,V)}{H(U,V)} + 1$$

Here $H(U,V)$ is the joint entropy and $H(U),\ H(V)$ are Shannon entropies of the two images U and V respectively. This quantity is used in lieu of the mutual information or cross correlation similarity measures because it is less sensitive to overlap [37] and hence facilitates comparisons across different tumors of differing information content.

For the synthetic data the true non-zero volume change percentages with respect to the reference volume of 42.19 cc was known from the design. The true, scaled homologous volumes ranged from 8.92 cc to 72.20 cc. The estimated vs true volume change percentages were compared using Bland-Altman plots [55] and linear regression.

3.3 Results

Figure 3-3 shows checkerboard plots for three registered tumors with alternating squares of the dilated manually segmented reference tumor and the "coffee-break" homologous mapped onto reference tumor volume. The overall good alignment can be noted from the figures.



(a) (b)

(c)

Figure 3-3 Checkerboard plots of dilated segmented reference tumor and homologous mapped onto reference tumor volume for 3 patients (a) Patient 1, Reference Tumor Volume = 10.86 cc (b) Patient 2, Total Reference Tumor Volume = 27.74 cc. Red ovals drawn to highlight the diffuse tumors (c) Patient 3, Reference Tumor Volume = 39.64 cc

Figures 3-4 and 3-5 show the mean percentage volume change error plots with respect to log of reference tumor volume for "coffee-break" data using mutual information and normalized cross correlation respectively. The 95% confidence intervals are depicted via the vertical lines.

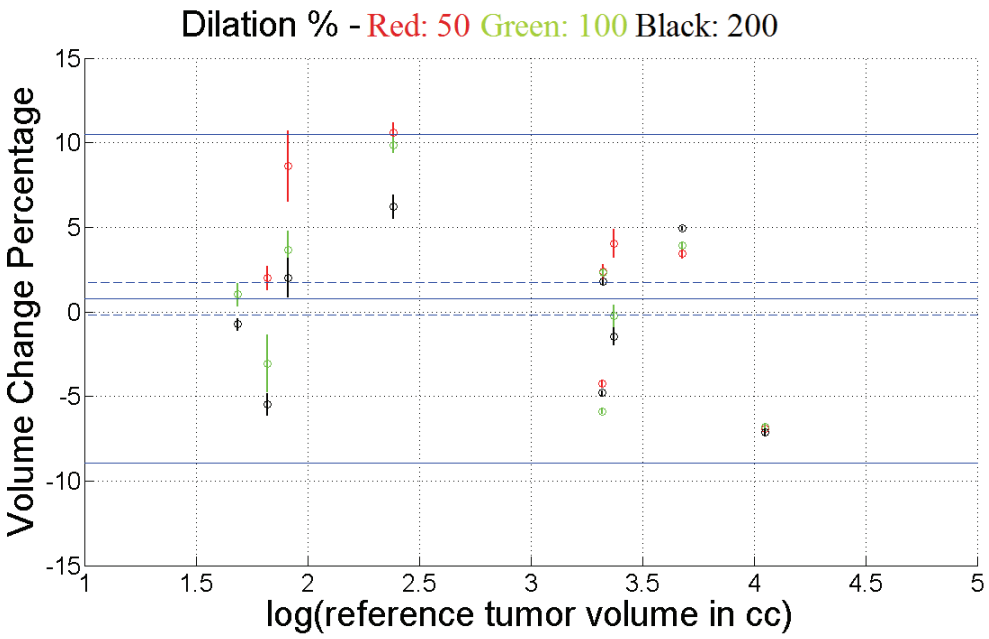


Figure 3-4 Percentage tumor volume change error plot with 95% confidence intervals for "coffee-break" data using *mutual information* as similarity measure. Red, green and black represent the three different dilations used. Mean error was 0.78% (95% CI: (-0.17%, 1.73%)), 95% limits of error were (-8.93%, 10.49%).

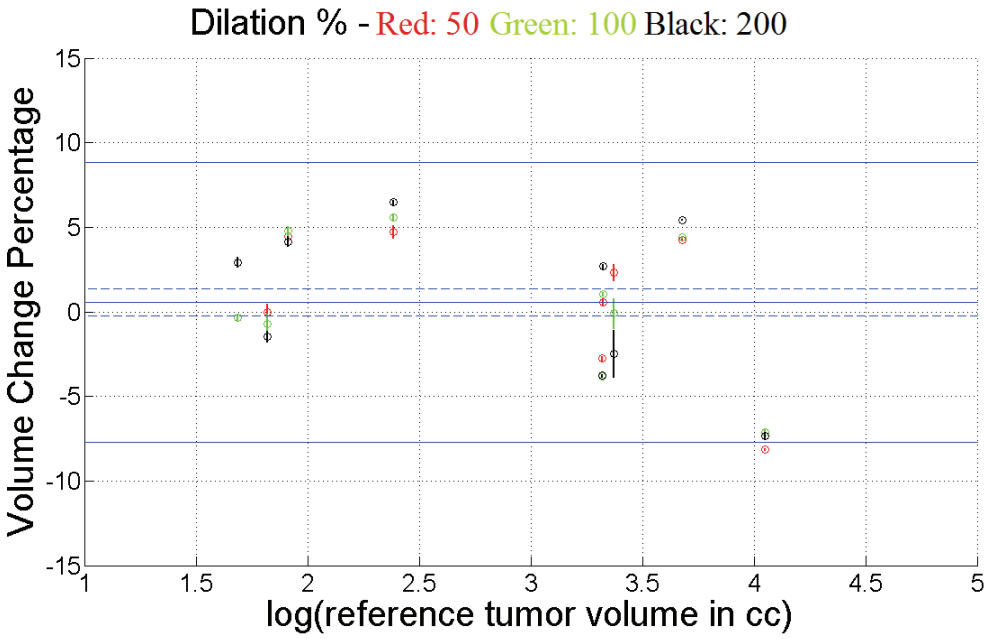


Figure 3-5 Percentage tumor volume change error plot with 95% confidence intervals for "coffee-break" data using *normalized cross correlation* as similarity measure. Red, green and black represent the three different dilations used. Mean error was 0.57% (95% CI: (-0.24%, 1.38%)), 95% limits of error were (-7.69%, 8.83%).

Table 3-1 and Table 3-2 show the parameter estimates of the fitted linear mixed-effects (lme) model for registrations using mutual information and normalized cross correlation respectively. The response variable was tumor volume change error. Table 3-3 and Table 3-4 show the parameter estimates of the fitted linear mixed-effects (lme) model for registrations using mutual information and normalized cross correlation respectively. The response variable in this case was percentage tumor volume change error. The similarity measure used for all models was the normalized mutual information of the registered tumor as mentioned previously.

Predictor	Name and units of Predictor	Estimate's 2.5 %ile	Estimate	Estimate's 97.5 %ile	Std. Error	p-value
$log\ V_r$	Log of reference volume in cc	0.6735	1.3199	1.9662	0.2803	0.0015
$S(U,V)$	Normalized mutual information (bits/bits)	-2.2847	-1.6629	-1.0412	0.3156	<0.0001

Table 3-1 Parameter Estimates of lme model for *volume change error* with 95% confidence intervals. Registration was done using *mutual information* as similarity measure. Normalized mutual information was used as the similarity measure in the lme model.

Predictor	Name and units of Predictor	Estimate's 2.5 %ile	Estimate	Estimate's 97.5 %ile	Std. Error	p-value
$log\ V_r$	Log of reference volume in cc	0.7208	1.4493	2.1779	0.3159	0.0018
$S(U,V)$	Normalized mutual information (bits/bits)	-2.6032	-1.9002	-1.1971	0.3569	<0.0001

Table 3-2 Parameter Estimates of lme model for *volume change error* with 95% confidence intervals. Registration was done using *normalized cross correlation* as similarity measure. Normalized mutual information was used as the similarity measure in the lme model.

Figure 3-6 and Figure 3-7 show the fitted 3D models for MI-based (Table 3-1) and NCC-based registrations (Table 3-2) respectively. The predictors are log of the annotated reference tumor volume in cc and the normalized mutual information of the registered tumor. The response is the log of volume change error i.e. the log of the unsigned difference between the estimated homologous and annotated reference tumor volumes. The points of intersection of the perpendicular black lines on the planes represent the coordinates of the main effects (i.e. normalized mutual information of the registered tumor and log of reference tumor volume in cc) obtained from the registrations.

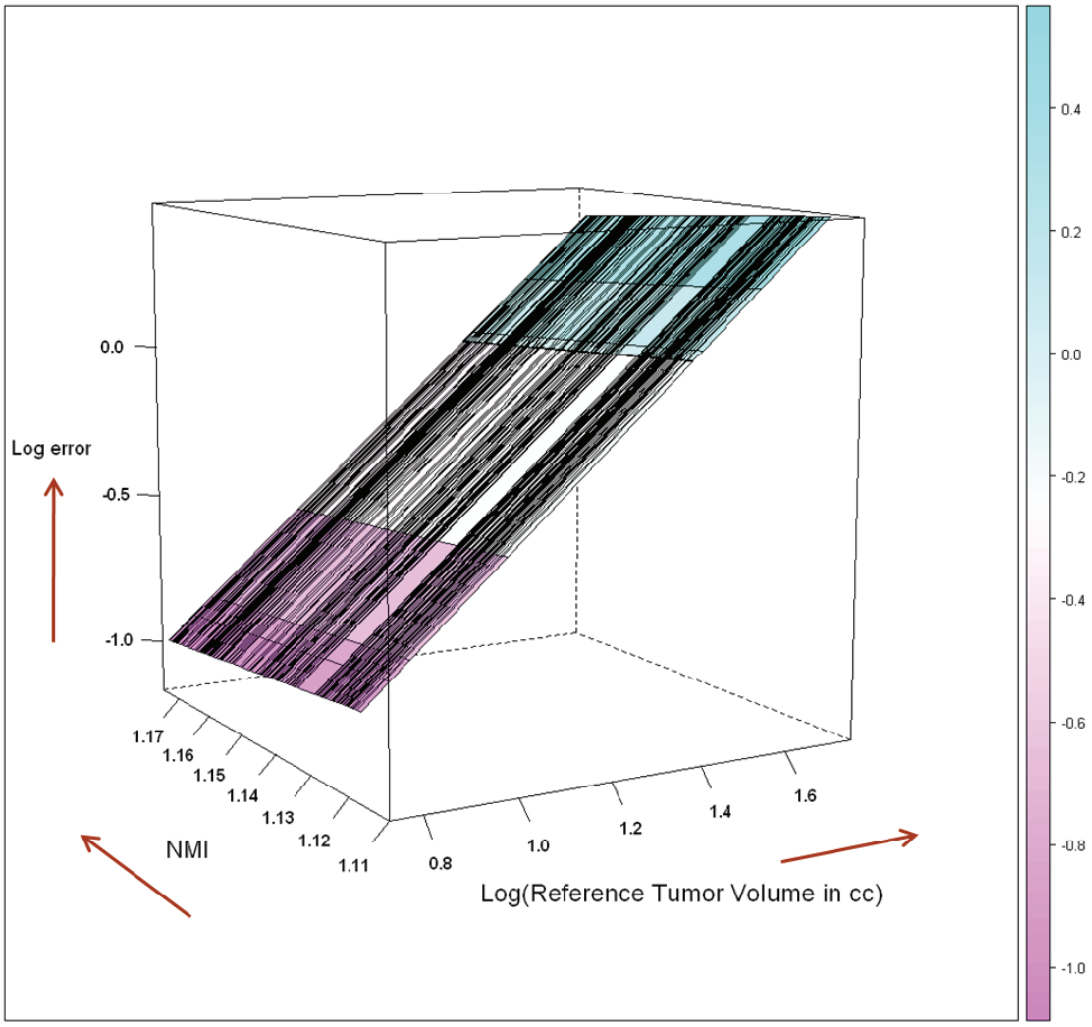


Figure 3-6 Predicted volume change error vs log of reference tumor volume and normalized mutual information for *mutual information-based registrations*. The points of intersection of the perpendicular black lines on the planes represent the coordinates of the main effects (i.e. normalized mutual information of the registered tumor and log of reference tumor volume in cc) obtained from the registrations.

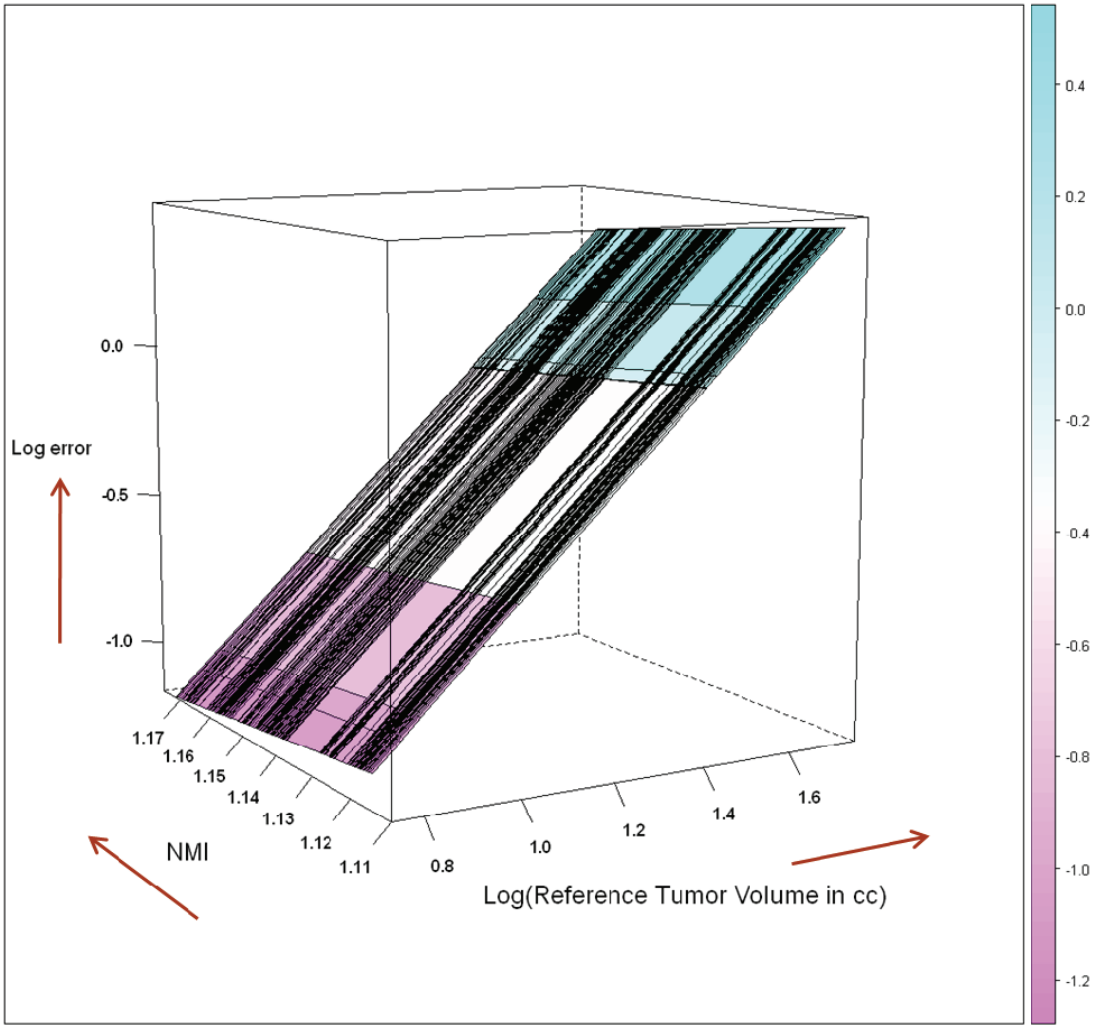


Figure 3-7 Predicted volume change error vs log of reference tumor volume and normalized mutual information for *normalized cross correlation-based registrations*. The points of intersection of the perpendicular black lines on the planes represent the coordinates of the main effects (i.e. normalized mutual information of the registered tumor and log of reference tumor volume in cc) obtained from the registrations.

Predictor	Name and units of Predictor	Estimate's 2.5 %ile	Estimate	Estimate's 97.5 %ile	Std. Error	p-value
$S(U,V)$	Normalized mutual information (bits/bits)	-136.8983	-83.3752	-29.8521	27.1811	0.0024

Table 3-3 Parameter Estimates of lme model for *percentage volume change error* with 95% confidence intervals. Registration was done using *mutual information* as similarity measure. Normalized mutual information was used as the similarity measure in the lme model.

Predictor	Name and units of Predictor	Estimate's 2.5 %ile	Estimate	Estimate's 97.5 %ile	Std. Error	p-value
$S(U,V)$	Normalized mutual information (bits/bits)	-71.9693	-40.5282	-9.0871	15.9670	0.0117

Table 3-4 Parameter Estimates of lme model for *percentage volume change error* with 95% confidence intervals. Registration was done using *normalized cross correlation* as similarity measure. Normalized mutual information was used as the similarity measure in the lme model.

For the synthetic exams with known non-zero volume change, the Bland Altman plot of the difference vs the mean of estimated and true volume change percentages is shown in Figure 3-8. 95% limits of agreement of the difference are included. The dashed blue line is the line of equality.

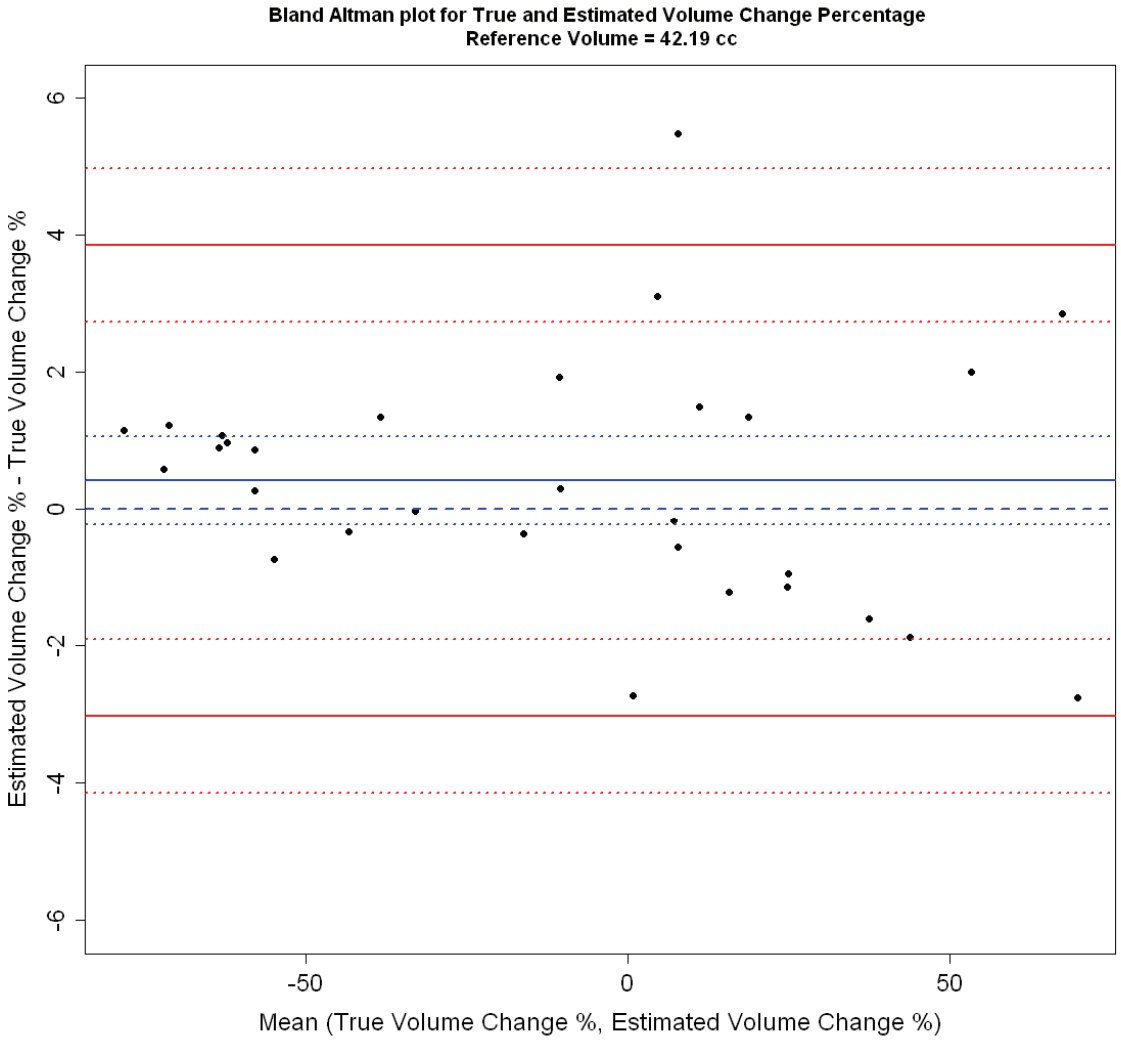


Figure 3-8 Bland Altman plot of the difference vs the mean of estimated and true volume change percentages. The solid blue line at 0.42% (95% CI: -0.23% to 1.07%) represents the mean difference. Note that the confidence intervals do not exclude zero bias. The 95% limits of agreement are shown with a solid red line. The lower and upper limits are -3.02% (95% CI: -4.14% to -1.90%) and 3.86% (95% CI: 2.74% to 4.98%) respectively.

Scatter plot of estimated vs true volume change percentages together with the mean and limits of agreement derived from the Bland Altman plot are presented in Figure 3-9. The regression parameters for estimated vs true volume change percentage are presented for comparison in Table 3-5 together with their 95% confidence intervals. The R^2 value was 0.9985.

Predictor	Estimate's 2.5 %ile	Estimate	Estimate's 97.5 %ile	Std. Error	p-value
Intercept	-0.3427	0.3364	1.0156	0.3316	0.3190
Slope	0.9779	0.9929	1.0079	0.0073	<0.0001

Table 3-5 Parameter Estimates of linear regression model for the synthetic exams (estimated vs true volume change percentage) with 95% confidence intervals.

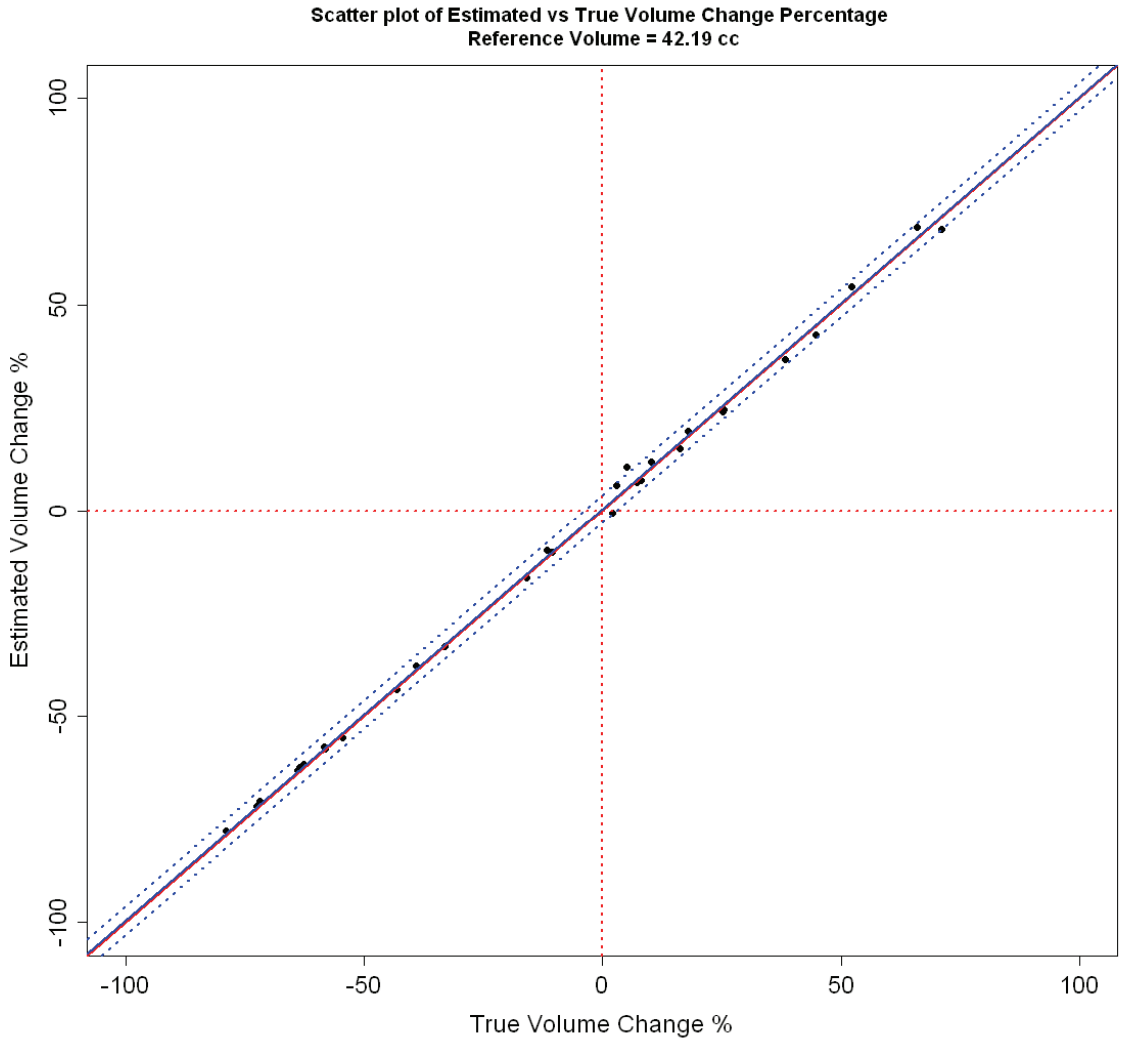


Figure 3-9 Estimated vs true volume change percentage for the synthetic exams. The mean bias and 95% limits of agreement derived from the Bland Altman plot are shown with blue lines. The diagonal red line is the line of equality.

3.4 Discussion

Accurate tumor volume change measurement is likely a reliable marker for the evaluation of the tumor's response to therapy. Most approaches to measure volume change involve sequential segmentation of the tumor in the two images. This study investigates the less explored approach of using image registration to quantify tumor volume change. The

good alignment visible on the checkerboard plots (Figure 3-3 (a) to (c)) suggests that the developed warping registration-based approach is capable of tracking the complicated shape evolution of tumors of varying size, morphology and geometry. Previous work to standardize tumor size change measurement and response set the cutoff bounds for response and disease progression using 1D RECIST and 2D WHO measurements to (-30%, 20%) and (-50%, 25%) respectively [9] [56]. Therasse et al., assuming spherical geometry, also suggested the bounds for 3D volume calculations to be (-65%, 73%) using RECIST and to be (-65%, 40%) using WHO [4]. Figure 3-4 and Figure 3-5 indicate that much tighter bounds in volume change calculations can potentially be obtained by using a registration-based algorithm for measuring tumor volume change than either RECIST or WHO. Comparison with segmentation-based approaches will be done in the next chapter. The mean volume change percentage error for the mutual information-based registrations was 0.78% with a 95% confidence interval of (-0.17%, 1.73%). The lower and upper limits for the 95% limits of the error were -8.93% and 10.49% respectively. The mean volume change percentage error for the normalized cross correlation-based registrations was 0.57% with a 95% confidence interval of (-0.24%, 1.38%). The lower and upper limits for the 95% limits of the error were -7.69% and 8.83% respectively. Note that zero is not excluded from the 95% confidence intervals for the bias in the mean error for both similarity measures. The volume change estimation is thus either unbiased across the population or the experiment is insufficiently powered to observe the bias. In this case more datasets would need to be examined in future to decrease the error on the 95% confidence limits for the mean bias.

The linear-mixed effects models fitted to the data (Table 3-1 and Table 3-2) indicate that the error in estimating tumor volume change is a function of both initial tumor volume and the normalized mutual information content of the tumor. The positive slope of the parameter estimates (1.32, p-value < 0.005 for MI and 1.45, p-value < 0.005 for NCC) associated with tumor size indicate that volume change error increases with increase in tumor size. This is consistent with the findings of Meyer et al. [50]. The negative slope of the parameter estimates associated with the similarity measure (-1.66, p-value < 0.0001 and -1.90, p-value < 0.0001) show that volume change error decreases with increase in the similarity measure across the population, i.e., increase in the normalized mutual information content of the tumor across the population. Figure 3-6 and Figure 3-7 graphically represent the results of Table 3-1 and Table 3-2 respectively and depict the same trends outlined above. For example, from Figure 3-6 for the MI-based registrations, an error of 0.1 cc for a 6.3 cc tumor can be expected to decrease to 0.08 cc when the normalized mutual information of the tumor goes up from 1.12 to 1.17. An error of 3 cc for a larger tumor of volume 63 cc can be expected to decrease to 2.5 cc for the same increase in normalized mutual information from 1.12 to 1.17. The lme models fitted to the absolute percentage volume change data (Table 3-3 and Table 3-4) show that percentage volume change error decreases with increase in normalized mutual information content of the tumor across the population (slope: -83.38, p-value < 0.01 for MI and slope: -40.53, p-value < 0.05 for NCC).

For the synthetic exams, estimated volume change percentage was compared to the true volume change percentage using a Bland Altman plot (Figure 3-8). The Bland Altman plot suggests a mean bias of 0.42% but the 95% confidence interval (-0.23%, 1.07%)

does not exclude zero. Note that zero is also not excluded (-0.34%, 1.02%) in the 95% confidence interval for the estimate of regression intercept as shown in Table 3-5. Since zero is not excluded in the 95% confidence interval, the volume change estimation is either unbiased or the experiment is insufficiently powered to detect the bias. The 95% limits of agreement for the population of differences between the estimated and true volume change percentages were (-3.02%, 3.86%) from the Bland Altman plot. The mean bias and 95% limits of agreement are low compared to the range of volume change being measured (-80%, 80%) as shown in Figure 3-8 and Figure 3-9. This indicates good agreement.

Chapter 4

Comparison of image registration-based and image segmentation-based tumor volume change measurement algorithms

4.1 Introduction

As mentioned previously, most efforts aimed at tumor volume change estimation are based on sequential segmentation. This is an indirect approach in that the tumor volume change between two interval scans is not obtained directly but via the subtraction of the independently segmented volumes in the two exams. This is in contrast to the registration-based approach described earlier wherein tumor volume change in a scan pair is measured directly. In this chapter we will look at a level set-based active contour segmentation approach for breast tumor volume change estimation to highlight the general challenges faced by a segmentation-based change measurement approach, discuss some of the segmentation approaches developed by other researchers for change analysis of pulmonary nodules and also compare some of these approaches with the proposed registration-based change measurement method.

4.2 Level set-based active contour segmentation of breast tumors

A sequential segmentation approach was used to measure the change in volume of breast tumors in the diffusion MRI "coffee-break" dataset described in the previous chapter. The level set-based 3D active contour segmentations were done using ITK-SNAP, a user friendly open source application software for 3D medical image segmentation [57]. Details of the implementation in ITK-SNAP can be found in [57]. Briefly, active contour evolution by region competition [58] was implemented using ITK sparse field level set algorithm. The contour is embedded as the zeroth level set [59] of a higher dimensional function ψ defined everywhere on the image domain and evolves according to the following partial differential equation:

$$\frac{\partial \psi(\vec{x}, t)}{\partial t} = F \nabla \psi$$

At any time t, the evolving contour can be obtained by solving for $\psi(\vec{x}, t) = 0$. F represents the sum of internal and external forces acting normally on the evolving contour and is shown in the equation below.

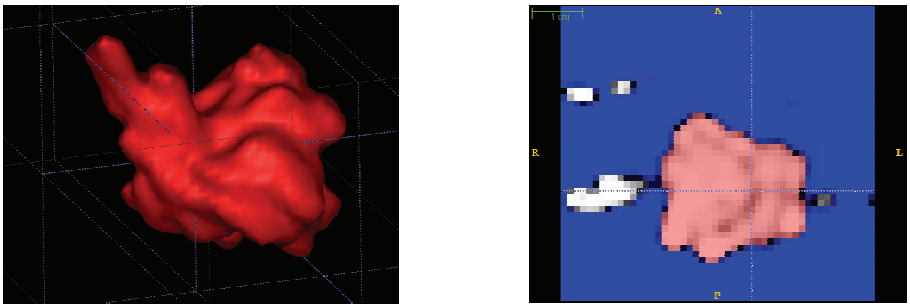
$$F = \alpha\left(P_{object} - P_{background}\right) + \beta\kappa$$

The internal force is proportional to the mean curvature of the contour κ and the external force is proportional to the difference of object and background probabilities obtained from a voxel probability map of the image domain.

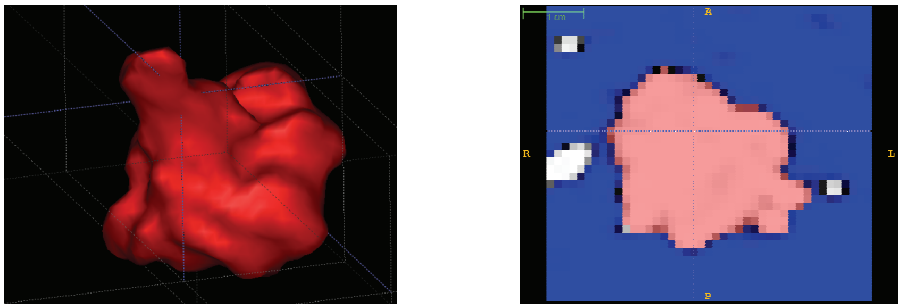
ITK-SNAP allows the user to choose appropriate values for the weighting parameters α and β that govern the evolution. α adjusts the propagation velocity and can be varied from -3 (contracting) to 3 (expanding) and β adjusts the curvature velocity and can be varied from 0.0 (detailed) to 0.5 (smooth) to 1.0 (spherical). A combinatorial approach was used for the parameters - varying α as 1, 2 and 3 and β as 0.1, 0.3 and 0.5 to span a range of feasible parameter values for the segmentations. Since all spherical seeds were placed entirely within the confines of the tumor, only expansive forces were chosen for α. Similarly the range of β values was chosen to reflect the irregular, complex shape of the breast tumors. The same intensity threshold range was chosen for both the scans of a given "coffee-break" pair except for one pair for which different thresholds had to be chosen due to high variations in intensity between the two "coffee-break" scans. The seeds were initialized in approximately corresponding or analogous locations determined visually in each exam of the "coffee-break" pair. This was done to minimize variability in the initialization process for a given pair of zero change exams. 8 out of the 9 "coffee-break" breast tumors were successfully segmented. One case (shown in Figure 3-3 (b)) could not be segmented using this algorithm as there were multiple small lesions spread throughout the breast and the segmentations leaked. Out of the 9 different combinations of α and β values, the values of α=2 and β=0.3 gave the best performance on average in terms of error in computing volume change (5.07%$\pm$22.20% S.D. of the average of the two computed tumor volumes) for the 8 tumors. Choosing the best (α, β) combination for an individual tumor in terms of error in percentage volume change improved the performance to 0.46%$\pm$16.42% S.D. of the average of the two computed volumes. At this point, however, it is best to refrain from drawing a comparison between the

performance of the registration-based algorithm and ITK-SNAP because ITK-SNAP is not optimized for measuring volume change. Comparison with the method proposed in this thesis and segmentation-based algorithms designed specifically for volume change measurement will be done in a later section. Instead this section will focus on the challenges faced during the level set segmentations which are in general true for any segmentation-based approach, i.e., a segmentation-based algorithm needs to address these challenges for accurate tumor volume change measurement.

The first challenge in a segmentation-based indirect change measurement algorithm is to ensure consistency (minimize the bias) of the segmentations across the two images – a difficult task due to variability caused by, among a host of other factors [60], image acquisition noise and patient motion. For example, uniform dilation or erosion by 1 mm of a lesion of diameter 23 mm increases or decreases its volume by 28% and -24% respectively. If two segmentations differ all around by 1 mm (about the size of a pixel) in segmenting this lesion, the volume change can be overestimated or underestimated by 28% or -24% respectively. Figure 4-1 shows a "coffee-break" breast lesion pair. True volume change is 0. Figure 4-1 (a) and (c) show the 3D segmented volumes on the two exams using the same segmentation parameters (i.e., thresholds, analogous seed point locations, propagation velocity weight and curvature velocity weight). Figure 4-1 (b) and (d) show one representative slice from each exam (not at the same scale). The relative difference in volume between the two zero change exams was computed to be 18.6%.



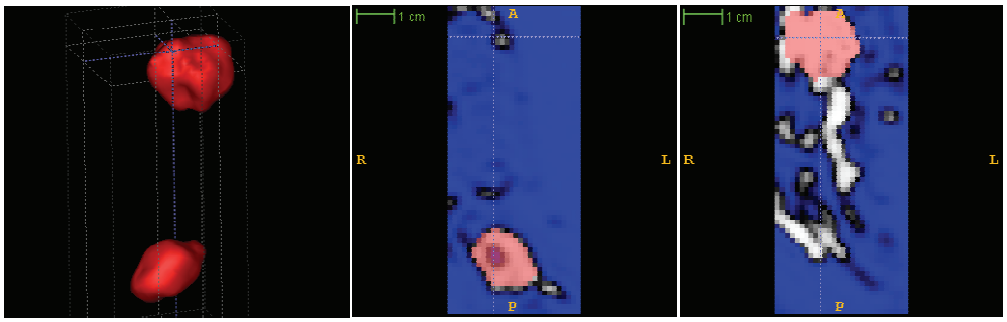
(a) (b)



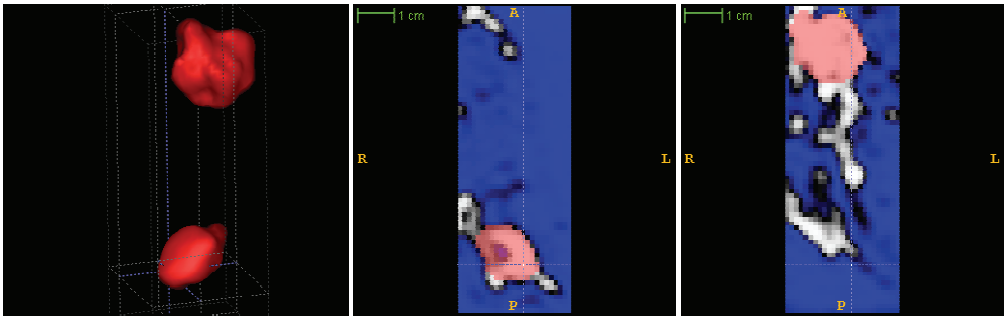
(c) (d)

Figure 4-1 Zero change breast lesion pair segmented with identical parameters. The relative difference in volume was 18.6% (a) segmented volume in exam 1 (b) one slice from exam 1 (c) segmented volume in exam 2 (d) one slice from exam 2

Figure 4-2 shows another zero change breast lesion pair. Figure 4-2 (a) and (d) show the 3D segmented volumes on the two exams using the same segmentation parameters (i.e., thresholds, analogous seed point locations, propagation velocity weight and curvature velocity weight). Figure 4-2 (b), (c) and Figure 4-2 (d), (e) show two representative slices from each exam. The relative difference in volume between the two "coffee break" exams was 8.8%.



(a) (b) (c)

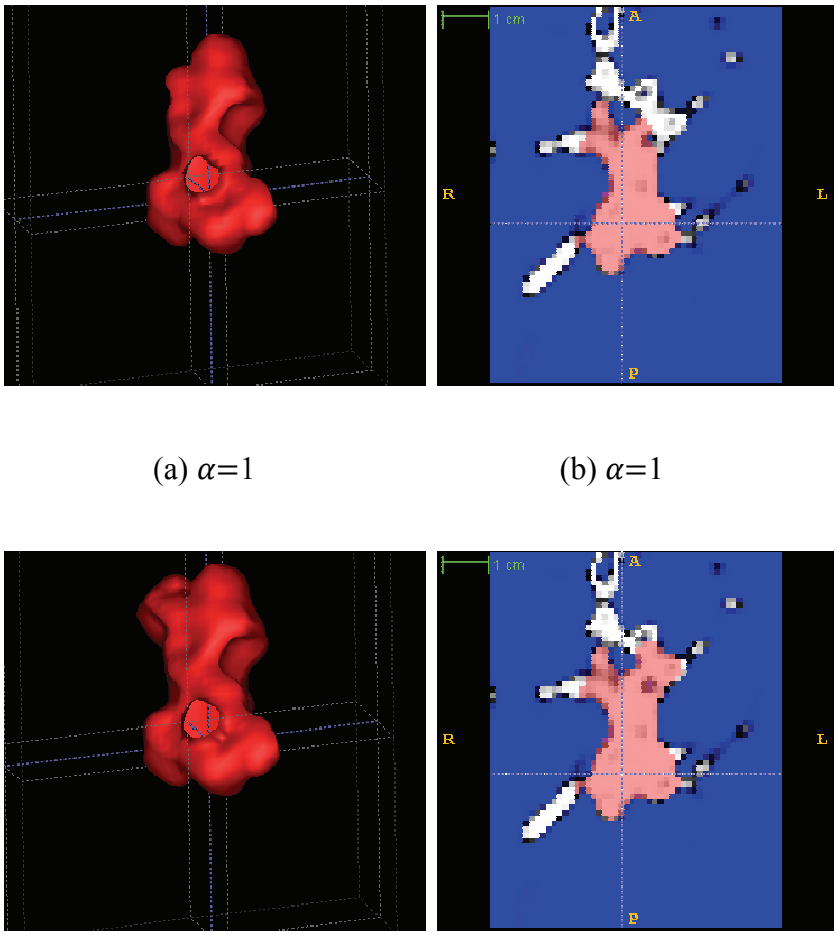


(d) (e) (f)

Figure 4-2 Zero change breast lesion pair segmented with identical parameters. The relative difference in volume was 8.8% (a) segmented volume in exam 1 (b) and (c) two slices from exam 1 (d) segmented volume in exam 2 (e) and (f) two slices from exam 2

The other major challenge in a segmentation-based change measurement algorithm is the issue of variability in *change* measurement. Changes in the employed segmentation parameters, say in the hands of different observers, can lead to widely different estimates of volume change for the same tumor. Inter-observer variability in segmentations can be quite high [50] and as variances add up, volume change calculations via subtraction of the two segmentations has twice the variance. Figure 4-3 and Figure 4-4 show the variability of the segmentations caused by varying segmentation parameters in the *same* scan. In Figure 4-3, all parameters in the scan (thresholds, seed point locations, curvature

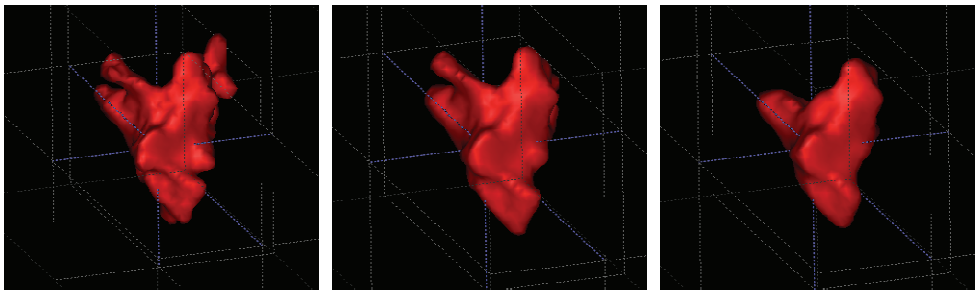
velocity weight) were kept same except for the propagation velocity weight α. The relative difference in volume between the two segmentations was 10.1%. In Figure 4-4, all parameters in the scan (thresholds, seed point locations, propagation velocity weight) were kept same except for the curvature velocity weight β. The relative difference in volume between the left and middle segmentations was 14.0% and between the left and right segmentations was 33.8%.



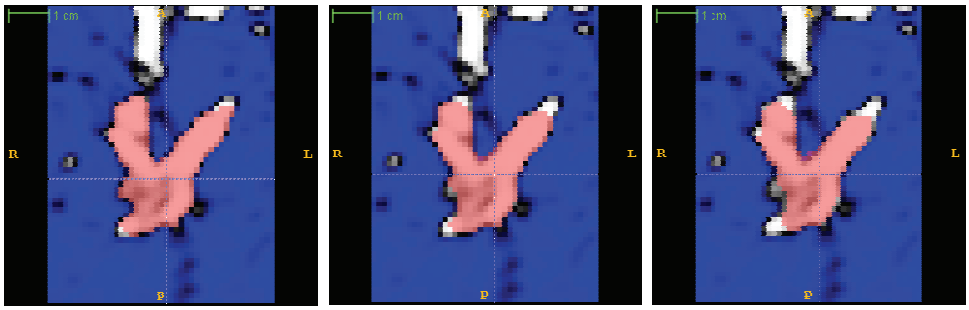
(a) α=1 (b) α=1

(c) α=2 (d) α=2

Figure 4-3 Variability due to change in segmentation parameters in the same exam. The relative difference in volume was 10.1% between the segmentations (a) segmented volume, α=1 (b) one slice from segmented volume, α=1 (c) segmented volume, α=2 (d) one slice from segmented volume, α=2



(a) β=0.1 (b) β=0.2 (c) β=0.3



(d) β=0.1 (e) β=0.3 (f) β=0.3

Figure 4-4 Variability due to change in segmentation parameters in the same exam. The relative difference in volume between the left and middle segmentations and left and right segmentations was 14.0% and 33.8% respectively (a) segmented volume, β=0.1 (b) segmented volume, β=0.2 (c) segmented volume, β=0.3 (d) one slice from segmented volume, β=0.1 (e) one slice from segmented volume, β=0.2 (f) one slice from segmented volume, β=0.3

4.3 Segmentation-based approaches for volume change measurement

Several segmentation-based volume change measurement algorithms are discussed in the literature largely aimed at pulmonary nodules [60]. As discussed in the previous section, volume change measurement accuracy depends on the accuracy of volume segmentations in the individual images as well as the consistency of the segmentations across the images and most efforts are directed towards improving these. Yankelevitz et al. [61], in an early study, used volumetric segmentation and an exponential model to determine the growth

rate of nodules. They also achieved a volume accuracy of ±3% for synthetic nodules. Okada et al. [62] used anisotropic Gaussian fitting and an ellipsoidal volume model for estimating the volumes of solid and nonsolid nodules. Goo et al. [63] studied synthetic lung nodules and recommended the use of statistically determined variable thresholds for volume segmentation. As in vivo lung nodules often have complicated attachments with vasculature, pleura or hilum, which also exhibit similar attenuation values, thresholding alone is rarely sufficient for successful segmentation. Kostis et al. [64] used supersampling and iterative morphological filtering to extract the lung nodule from the pleural surface and attached vessels. They reported successful segmentations, as determined by a radiologist, without manual intervention in 80% of nodules with vascular attachments and 72% nodules with pleural attachments. Kuhnigk et al. [20] also used morphological filtering but adaptively determined the morphological opening parameters instead of using fixed parameters like Kostis et al. In the final step of volume determination, they improved on simple voxel counting methods by computing partial voxels as well. In a study of "coffee-break" scans, they reported an absolute median percentage error of 4.7% with a 95% error limit of 26.9% for 96 out of 105 nodules. Volume change measurements could not be made for 9 out of 105 nodules due to leakage onto adjoining structures. A partial volume analysis approach was also recommended by Ko et al. in their study of synthetic lung nodules [65]. The Lesion Sizing Toolkit [66] developed by Kitware used a shape detection level set for nodule segmentation and for 6 "coffee-break" scans reported a mean error of 7% with a standard deviation of 9%. In another "coffee-break" study conducted on lung nodules isolated in lung parenchyma, Gietema et al. [67] reported a mean difference volume of 1.3% with (-21.2%, 23.8%) as

the 95% limits of agreement. Siemens LungCare was used for the segmentation of the nodules. LungCare was also used for volume change measurement of 151 zero-change nodules by Wormanns et al. [19] and they reported their mean volume change error and 95% limits of agreement as 0.7% and (-20.4%, 21.9%) respectively. Reeves et al. [46] used adaptive thresholding, 3D rigid body registration followed by rule based segmentation adjustments to improve the consistency of volume change measurements across intervals. In a study of 50 stable (diagnosed as benign via biopsy or observed to have no growth over two years) nodules they were able to see a reduction in the standard deviation of percent volume change calculations from 11.54% to 9.35% ($p=0.03$) using this method as opposed to employing independent fixed threshold segmentations. Jirapatnakul et al. [68] reported a mean percentage volume change of 1.98% and (-25.2%, 18.6%) as the 95% limits of agreement for semi-automated segmentation of zero-change datasets with same slice thickness for both images. When scans of differing slice thicknesses were included the mean percentage error and the 95% limits of agreement rose to -12.2% and (-54.6%, 30.3%) respectively. The VOLume Change Analysis of Nodules (VOLCANO) 2009 challenge had 13 teams using 17 different methods for the change analysis of 50 lung nodules [69]. The results are summarized in the proceedings of the Second International Workshop on Pulmonary Image Analysis [70].

4.4 Comparison with registration-based approach

Publicly available datasets with known truth are essential for objective comparison of different change measurement methodologies. As part of the National Cancer Institute exercised Reference Image Database to Evaluate Therapy Response (RIDER) contracts [18], two such sources – United States Food and Drug Administration (FDA) phantom

data and Memorial Sloan-Kettering Cancer Center (MSKCC) “coffee-break” data – are available via the National Biomedical Imaging Archive website. The registration-based change measurement algorithm discussed in the previous chapter was tested on part of the FDA phantom data and the whole of the MSKCC zero change data. Krishnan et al. [66] have used the Lesion Sizing Toolkit to segment and measure nodule volumes on an identical subset of the FDA dataset. Zhao et al. [71] have evaluated the variability of segmentation-based tumor volumetric measurements in repeat CT scans using the same MSKCC data.

4.4.1 FDA phantom data

The FDA has made publicly available CT scans of an anthropomorphic thoracic phantom with vasculature and synthetic nodules of varying sizes attached to the vasculature [72]. The nodules come in four different shapes (spherical, elliptical, lobulated, spiculated), six different sizes (diameters ranging from approximately 5 mm to 40 mm) and four different densities (ranging from -800 HU to 100 HU). In a combinatorial approach, different scans were acquired by varying dose, pitch and slice collimation and then reconstructed by varying slice thickness and reconstruction kernels. The phantom was repositioned for each new imaging protocol. For each of the scanned nodules, true volumes were determined from weights measured via a precision scale of 0.1mg tolerance and the manufacturer provided densities.

Six different CT scans were analyzed from among those available using the registration-based change measurement algorithm introduced previously. These scans were chosen to be consistent with the ones analyzed with the Lesion Sizing Toolkit by Krishnan et al. [66] so that a comparison between the algorithms could be made. Each scan had an

approximately 5 mm, 8 mm and 10 mm diameter spherical nodule of density 100 HU attached to vasculature. Three of the scans had slice thicknesses of 0.8 mm and slice increments of 0.4 mm and the other three had slice thicknesses of 3 mm and slice increments of 1.5 mm. Two scans had a dose of 25 mAs, two had 100 mAs and the remaining two had 200 mAs. All scans were acquired using a Philips 16-row multidetector scanner Mx8000 IDT and slice collimation of 16x0.75 mm and pitch 0.9 and reconstructed using a detailed reconstruction kernel. Thus effectively for each of the three nodules there were 6 “coffee-break" scans with zero volume change. However variations in acquisition (dose) and reconstruction (slice thickness) parameters together with the noise attached to phantom repositioning for each protocol is expected to make volume change measurements noisy.

The first step in the registration-based algorithm was to obtain a mask of the nodule in the reference image. For each of the nodules, the 200 mAs scan with slice thickness 3 mm and slice increment 1.5 mm was chosen as the reference. Level set-based active contour segmentations with region competition were performed to obtain the nodule masks using ITK-SNAP [57]. The estimated reference volumes using ITK-SNAP and the true volumes provided by the FDA are compared in the table below.

Nodule approximate diameter (mm)	True volume (µl)	Estimated reference volume (µl)	Error in estimation (% of true volume)
5	64	60.23	-5.89
8	255	299.33	17.38
10	506	538.43	6.41

Table 4-1 FDA phantom data: true vs estimated volumes

Before registration, the reference volume was dilated using 5 iterations of a 3x3 binary structuring element. A set of hexagonally close-packed control points were automatically generated in the 3D reference volume spaced 6 mm apart. This spacing was chosen so that at least 5 control points (minimum required to compute a thin plate spline warp) were deployed on the dilated reference volume for the smallest nodule (5 mm diameter). This low degree of freedom warping approach is sufficient for registering these largely homogeneous nodules. The rest of the algorithm used for registration is similar to the one described for the registration of breast tumors in Chapter 3. Five registrations were done per nodule between the reference image and the other 5 "coffee-break" homologous images obtained from varying exposure and slice thickness. A total of 15 registrations were performed (five/nodule). The mean volume change percentage error for the three nodules was 1.19% with a standard deviation of 5.60%. For the 5 mm diameter nodule the mean error was 2.09% with a standard deviation of 4.65%. The corresponding statistics for the 8 mm diameter and 10 mm diameter nodules were 2.18%$\pm$8.02% and -0.72%$\pm$4.12% respectively.

As mentioned previously, Krishnan et al. used the Lesion Sizing Toolkit to measure nodule volumes on an identical subset of the FDA phantom data. The Lesion Sizing Toolkit follows a modular approach to segmentation. Separate feature generator modules are used in a region of interest defined tightly around the tumor to detect strong edges, detect the lung wall, detect surrounding vessels and threshold intensities. A feature aggregator module combines the information from these generated features to form an aggregated feature image. A shape detection level set is then implemented and the contour evolution proceeds using the aggregated feature image as the propagation

component and the contour's curvature as an internal force component. For all three nodules, they reported a mean error of 35% with respect to the true volume provided by the FDA. Volume change was not computed but the standard deviation of the volume measurements was reported as 20% of the true volume. These being zero change nodules any deviation from zero can be viewed as error. Since variances add, the standard deviation or error in volume change measurements can be expected to be $\approx 20\sqrt{2}\% = 28\%$ of the true nodule volume. Compare this with the standard deviation of 5.60% of the estimated reference nodule volume for the estimates obtained from the registration-based method. Since the segmentations had a mean bias of 5.97% (from Table 4-1) of true nodule volume, this implies that the standard deviation or error in the registration-based volume change measurements was 5.60% of the estimated volume or 5.93% of true nodule volume.

4.4.2 MSKCC "coffee-break" data

The Memorial Sloan-Kettering Cancer Center, New York acquired same day repeat CT scans of 32 patients aged 29 to 82 years (mean age 62.1 years) [71]. All patients were diagnosed with non-small cell lung cancer. Both scans for a patient were acquired under breath hold within 15 minutes of each other and the patients were required to leave the scanner table after the first scan. Scans were acquired using a 16 detector row (GE LightSpeed 16) or 64 detector row (GE VCT) CT scanner. While the in-plane resolutions varied, the slice thickness of all scans was 1.25mm. As described in Zhao et al. [71], the nodules had complicated geometry and 18 out of the 32 nodules were attached to one or many surrounding anatomic structures including the mediastinum, hilum, pleura and chest wall.

The registration-based change measurement algorithm discussed in the previous chapter was used for the volume change quantification of these nodules. For each of the 32 patients one CT scan was randomly chosen as the reference and the other exam from the "coffee-break" pair served as the homologous. The tumors were segmented only in the reference exams by implementing level set active contour segmentations with region competition in ITK-SNAP. Manual correction, if required, was used to prevent any leakage into the surrounding vasculature, pleura or hilum. Please note that the manual correction was applied *before* the registration process in order to capture the extents of the tumor as accurately as possible. The registration-based algorithm did not use *multiple* segmentations or annotations to determine volume change. The segmentation was in the reference volume only and was used as a region of support for the registration (after dilation) and subsequent integration of the Jacobian of the deformation for the volume change computation. The segmented reference was dilated by 5 iterations of a 3x3 binary structuring element. Around 8 to 30 control points were distributed in a hexagonal close packed arrangement in the dilated reference volume depending on the size of the nodule, i.e., the number of control points deployed increased as the size of the nodule increased. As the nodule volumes varied widely in size (from 1 cc to 150 cc), no one density of hexagonal close packing was found to be optimal. Instead hexagonal close packed control point spacings were varied between 8 mm to 22 mm to generate 8 to 30 control points on the reference volumes. It was found that a relatively low number of control points were sufficient to accurately warp lung nodules. This is in contrast to the breast tumors discussed in Chapter 3 which had a much higher mutual information density/unit volume and could support higher degrees of freedom. The nodules were registered using a

multiscale thin plate spline warping registration algorithm as described in Chapter 3. The results of the volume change percentage error for these zero change nodules are shown in Figure 4-5. The mean error was 2.55% with a 95% confidence interval of (0.08%, 5.03%). The 95% limits of agreement were -10.90% (95% CI: -15.14% to -6.67%) to 16.01% (95% CI: 11.78% to 20.25%). As zero is excluded from the 95% confidence interval of the mean error, there is a possibility that the volume change estimation has a small positive bias.

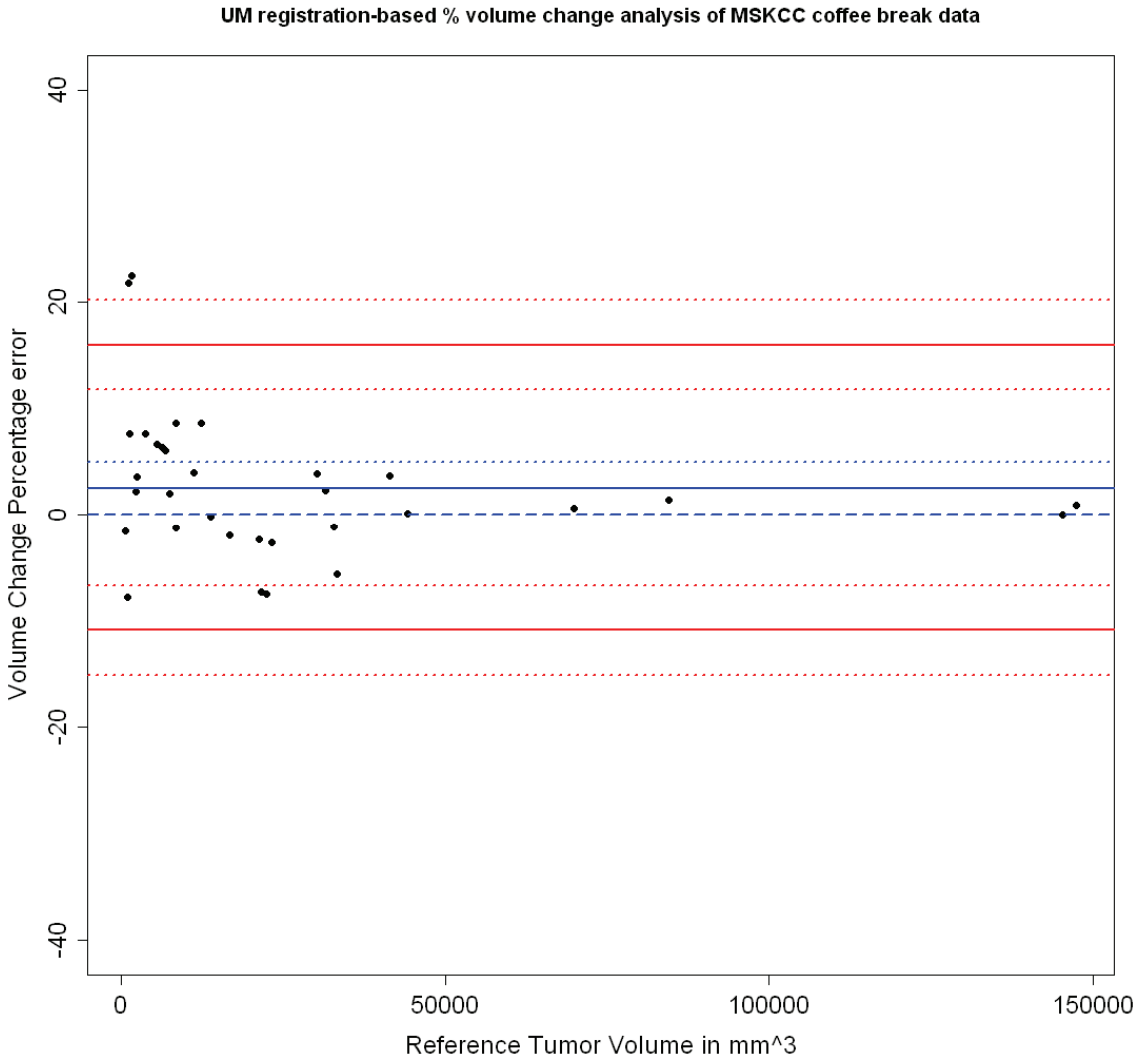


Figure 4-5 Plot of volume change percentage error vs reference tumor volume for the zero change MSKCC data. The mean error was 2.55% (95% CI: 0.08% to 5.03%) with 95% limits of agreement between -10.90% (95% CI: -15.14% to 6.67%) to 16.01% (95% CI: 11.78% to 20.25%). Excluding the outliers the mean error was 1.25% (95% CI: -0.5% to 3.0%) with 95% limits of agreement (-7.96%, 10.46%)

However, this bias is largely due to the two outliers visible on the top left of the plot (Figure 4-5). Both of these were small, noisy nodules and it is likely that the registration did not do as well because of the small and noisy clusters in the bivariate histogram. Excluding these nodules, the mean error was 1.25% with a 95% confidence interval of

(-0.5%, 3.0%) and the 95% limits of agreement were (-7.96%, 10.46%). Since zero is not excluded in the 95% confidence interval for the mean error, it is possible that the volume change estimation is, in general, unbiased over the whole gamut of lung nodules of varying shapes and sizes. More small, noisy nodules would need to be analyzed in future to ascertain if there is a systematic positive bias associated with such nodules or if the large positive errors in this case were random and simply a result of registration error.

Zhao et al. used the semi automatic segmentation algorithm described in [71] to segment the nodules in the 32 "coffee-break" scan pairs. Briefly, the algorithm implements an iteratively decreasing threshold level segmentation in a region of interest that tightly encloses the nodule. At each threshold level, the largest geometrically connected contour is identified and the threshold level that provides the maximum surface gradient is chosen as optimal for the segmentation. They computed the mean relative difference of the tumor volumes on the average of the two tumor volumes. They obtained a mean relative error of 0.7% and (-12.1%, 13.4%) 95% limits of agreement from a Bland Altman plot analysis. However, expert radiologists considered 13 out of the 32 nodules (41%) as unsatisfactorily segmented due to leakage onto adjoining structures. Segmentations for these scan pairs were manually corrected side by side by an expert radiologist before computing the volume change. While manual side by side correction in both scans worked out well in the case of these 32 "coffee-break" scan pairs, based on previous findings of other researchers, we feel that such an approach can prove suboptimal in many cases, especially when the nodules change in shape or size. Manual annotations, even when done by expert radiologists, have been shown to be a major source of variability [50]. Reeves et al. [73] showed that the limits of agreement among radiologists

for a 10 mm nodule were between (-23.4%, 27.7%). Also, since variances add, volume change calculations based on the difference of two segmentations is expected to double the variability of the results.

4.5 Discussion

Most segmentation algorithms measure change indirectly. Segmentations are done independently in the two images and volume change is computed via subtraction of the segmented image volumes. This independent approach ignores crucial mutual information present in the two images of the tumor. Also variances add, making change measurements based on volume subtractions more noisy.

The registration-based algorithm discussed previously does not use multiple segmentations to compute change. Only one segmentation is done in the reference image to provide a region of support for the registration and subsequent integration of the Jacobian of the deformation. While accurate segmentation in the reference image is desired for accurate computation of percentage tumor volume change, a registration-based change measurement algorithm is relatively insensitive to small errors in this segmentation. This is because before registration the segmented reference mask obtained automatically, semi automatically or via expert annotation, is dilated to include information from the tumor surroundings which helps the registration of the tumor in the two scans. For example, in the case of a lung nodule attached to multiple vessels and the pleura, the entire nodule and its surroundings are registered due to the dilation of the segmented mask. If the nodule changes size (due to progression in disease or response to therapy) and the vessels and the pleura do not, registration will reflect the result in the

Jacobian values of the computed deformation. The Jacobian values will be 1 (or close to 1) on the vessels and pleura indicating little or no change in size and greater than or less than 1 on the nodule depending on the growth or contraction of the nodule volume. If the volume change is calculated over this entire region it will be accurate in that it will reflect that only the nodule changed size. The computation of the volume change percentage (calculated over the segmented reference tumor), on the other hand, will depend on the accuracy of the segmentation or annotation and as long as that is accurate to a reasonable degree of tolerance, true or close to true volume change percentage of the nodule can be computed. On the other hand for segmentation-based change measurement methods, if the segmentations of the object of interest end up being different or inconsistent across scans, say due to variability associated with image noise or change in shape, size or attenuation of the tumor across intervals, both the calculated volume change and volume change percentage (calculated over the average of the two segmented volumes) will be erroneous. Changes in nodule morphology or its degree of attachment to surrounding anatomical structures can adversely affect the accuracy and consistency of segmentations. However, changes in shape and size are easily captured by an accurate warping registration algorithm. Also interval changes or variations in attenuation in the nodule simply show up as different clusters in a mutual information-based registration algorithm's joint density function and the accuracy of the registration (and thus the volume change measurement) is not adversely affected. In addition, a registration-based change measurement algorithm can also detect differential growth or contractions across tumor sub-volumes which segmentation-based approaches generally ignore.

Chapter 5

Tracking chronic obstructive pulmonary disease using image registration-based volume change measurements: an initial study

5.1 Introduction

In the previous chapters an image registration-based algorithm for accurate quantification of tumor volume change has been developed. The tools and concepts discussed, however, are not limited to the estimation of volume change in tumors. Another potential application of the image registration-based volume change quantification algorithm is in tracking the progress of obstructive lung disease in patients. Chronic obstructive pulmonary disease or COPD is a progressive disease of the lungs that impairs normal lung function. COPD mainly refers to the presence of one or both of the following conditions – emphysema, a condition in which the alveoli lose their functional shape making gas exchange difficult and chronic bronchitis, a condition characterized by inflammation of bronchi that makes it harder to breathe. COPD is diagnosed clinically via spirometry tests and the FEV_1/FVC ratio of the lungs is used to determine the stage of the disease [74].

Image registration has been used to track rates of lung expansion and contraction and subsequently compared with spirometry data [75]. Reinhardt et al. used 3D inverse

consistent image registration [76] using sum of squared differences as the similarity measure and compared the degree of regional lung expansion measured by the Jacobian of the deformation to xenon CT based measures of specific ventilation in five anesthetized sheep [77]. Even though the correlation coefficients of the linear regression between the average Jacobian and average specific ventilation varied widely among the five animals (0.08 to 0.94), a linear relationship between the two was apparent. Li et al. [78] used an intensity and landmark-based inverse consistent image registration algorithm to track lung tissue changes during the respiratory cycle. In the studied human subject, they demonstrated good global alignment of the lung volumes including lobar fissures. Some misregistration, however, is visible in the alignment of the internal features of the lung. To track lung tissue changes accurately during respiration, good alignment in both global and local features is necessary. In this chapter mutual information-based 3D warping image registration will be used to accurately register and track volume changes in the lung tissues of patients diagnosed with chronic obstructive lung disease. Changes in global and regional lung volume are lower in diseased lungs than normal lungs. The Jacobian values of the deformation obtained from registration can, therefore, potentially be used to quantify the state of the obstructive disease.

This chapter is exploratory in nature and intended as a proof of concept. In future, several patients with varying degrees of obstructive lung disease would need to be analyzed to see if registration can be used to track progress of COPD in longitudinal studies.

5.2 Methods and Materials

Patients who underwent diagnostic CT scans were graded 1 through 4 based on clinical exam reports. Patients with moderate to severe disease were classified as Grade 1, mild to moderate disease patients as Grade 2, the ones with mild disease as Grade 3 and the "normal" patients as Grade 4. Testing was done on two extreme datasets (one case from Grade 1 and one from Grade 4) to see if the Jacobian of the deformation computed from the registration was sensitive to the stage of lung disease. For example, a patient with severe obstructive disease can be expected to exhibit relatively lower increases in lung volume during full inspiration than a normal person. It was expected that this difference will be apparent in the Jacobian values of the computed deformation.

In this section the process used to register the lungs during different phases of the respiratory cycle in patients with COPD will be described. Two CT scans of each patient's thorax were acquired during expiration phase and inspiration phase using a GE LightSpeed16 scanner. The scans were acquired using 120 kVp, 100 mAs with an in-plane resolution of 0.74 mm and slice thickness of 1.25 mm. The slice spacings in the expiration scan were 20 mm and in the inspiration scan they were 1.25 mm.

5.2.1 Segmentation

The human lungs reside within the thoracic cavity with a slip boundary separating them from the thoracic wall. Due to their varied elasticity, the deformations of the lungs, the heart and the thoracic wall can differ considerably from each other during a respiratory cycle. The presence of the chest wall and heart can therefore hamper the registration of the lungs by dramatically increasing the degrees of freedom needed to describe the

deformation. Hence a 3D region growing segmentation algorithm was implemented, using ITK modules [79], as a preprocessing step to extract the two lungs from the expiration CT scan. This ensured that the focus of the registrations remained on the lungs and also obviated the need to deploy additional control points on the periphery of the lungs to track the movement of the chest wall and heart in unison with that of the lungs.

Two seeds are required for initializing the 3D region growing segmentation process. The first seed is used to extract regions containing the chest wall and the heart. The second seed is used to extract regions containing the DICOM padded values (-2000) and the regions beyond the chest wall. Thresholds for the region growing process are selected from volumes of interest placed on the regions to be extracted. The results of these two region growing segmentations are combined and subtracted from the original image volume to obtain the segmented lungs.

5.2.2 Registration

All registrations were done using MIAMI Fuse (Mutual Information for Automatic Multimodality Image Fusion) [36], a software suite for the automatic alignment of medical datasets, built on Advanced Visualization System 5 (AVS). Registrations were performed to map the geometry of the patient's lungs in the expiration phase into the geometry of their inspiration phase. The densely sampled inspiration field (spaced 1.25 mm apart) was always chosen as the homologous image during the registration as it was interpolated onto the frame of the reference image. The sparsely sampled segmented expiration image (with slices spaced 20 mm apart) was chosen as the reference. The segmented reference was dilated by 2 iterations of a rectangular 3x3 binary structuring element to help capture edge information. A multiscale registration algorithm, similar to

that described in the previous chapters, was used for the thin plate spline warping registrations. Mutual information was chosen as the similarity measure as it is more robust than the sum of squared differences to the variations of HU in the lung parenchyma caused by the entry and exit of air through the lungs.

Four hundred distance sorted control points (around 20/lung slice on average) were placed uniformly in the reference volume for the Grade 1 patient during the registration. For the Grade 4 "normal" patient, multiple registrations were performed 2 to 3 slices at a time and 1 lung at a time. Breaking up the whole registration process for the two lungs into multiple smaller registrations of 2-3 slices/lung allowed the registration to proceed with locally higher degrees of freedom required for the accurate registration of these high information datasets. Throughput was increased by doing the multiple registrations in parallel. Decoupling the large movements of the patient's left and right lungs also makes it easier for the thin plate splines to compute the deformation. It should be noted that even though these lung datasets require relatively high degrees of freedom for accurate registration, the degrees of freedom required is highest for Grade 4 patients (around 40 control points/lung slice on average) and lowest for Grade 1 patients due to the stiffness of the diseased lung.

5.3 Results and Discussion

Checkerboard plots from representative slices for both patients are included in Figure 5-1 and Figure 5-2. Figure 5-1 shows checkerboard plots of the expiration vs. the inspiration mapped onto the expiration image for the Grade 1 patient clinically classified as having moderate to severe disease. Figure 5-2 shows similar checkerboard plots for the Grade 4

patient clinically classified as “normal”. Both the Grade 1 and Grade 4 cases were successfully registered – the Grade 4 patient requiring higher degrees of freedom to register than the Grade 1 patient. The complexity of registering Grade 2 and Grade 3 patients is expected to lie somewhere between these two grades. Grade 1 patients with severe obstructive disease typically exhibit small deformations between inspiration and expiration respiratory cycles and the transformation is easily captured using the deformation interpolant (in this case thin plate splines). Grade 4 or “normal patients” typically exhibit much larger and more complex deformations between respiratory cycles which are more difficult to capture and require higher degrees of freedom.

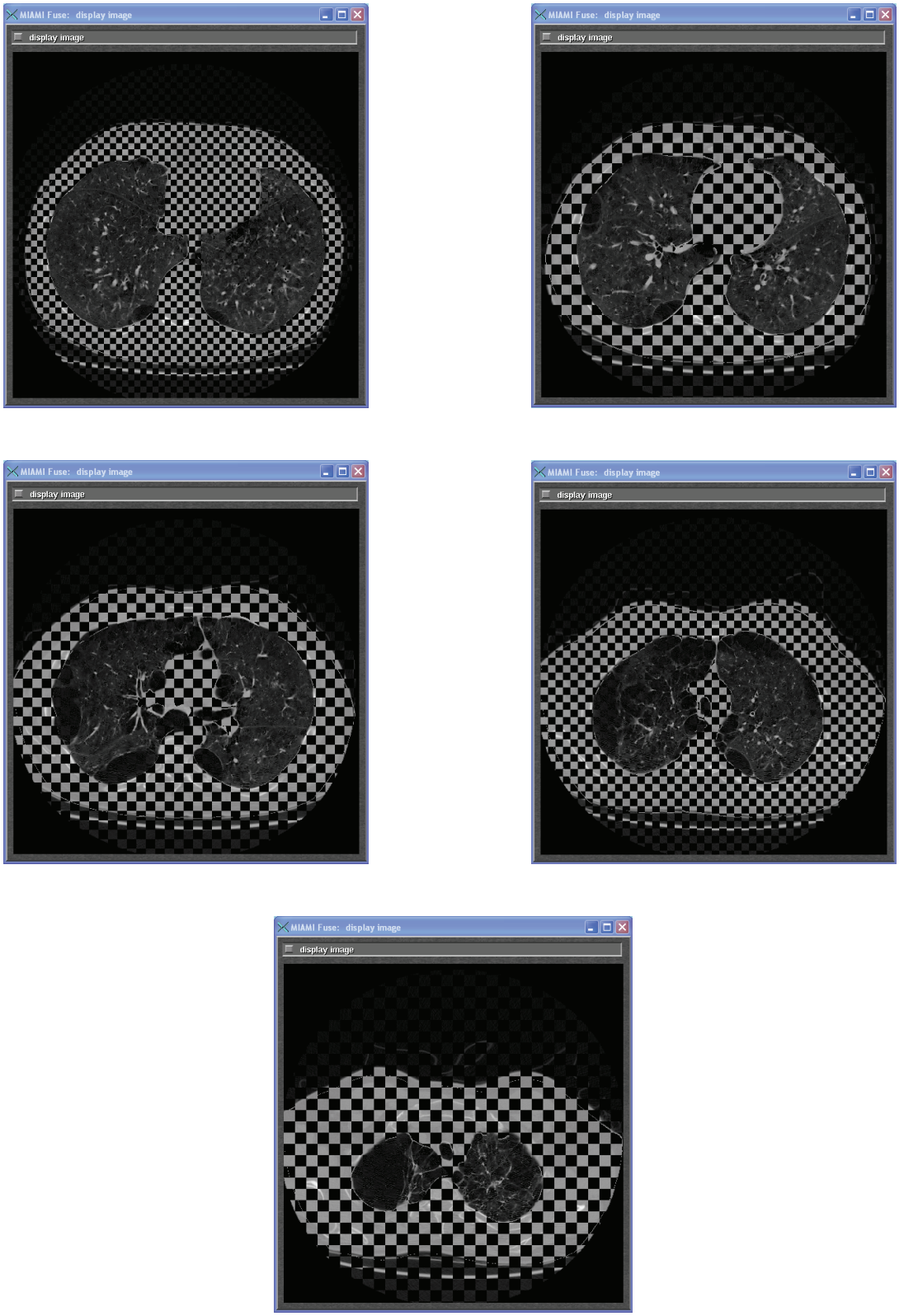


Figure 5-1 Checkerboard plots of registration for patient with moderate to severe disease

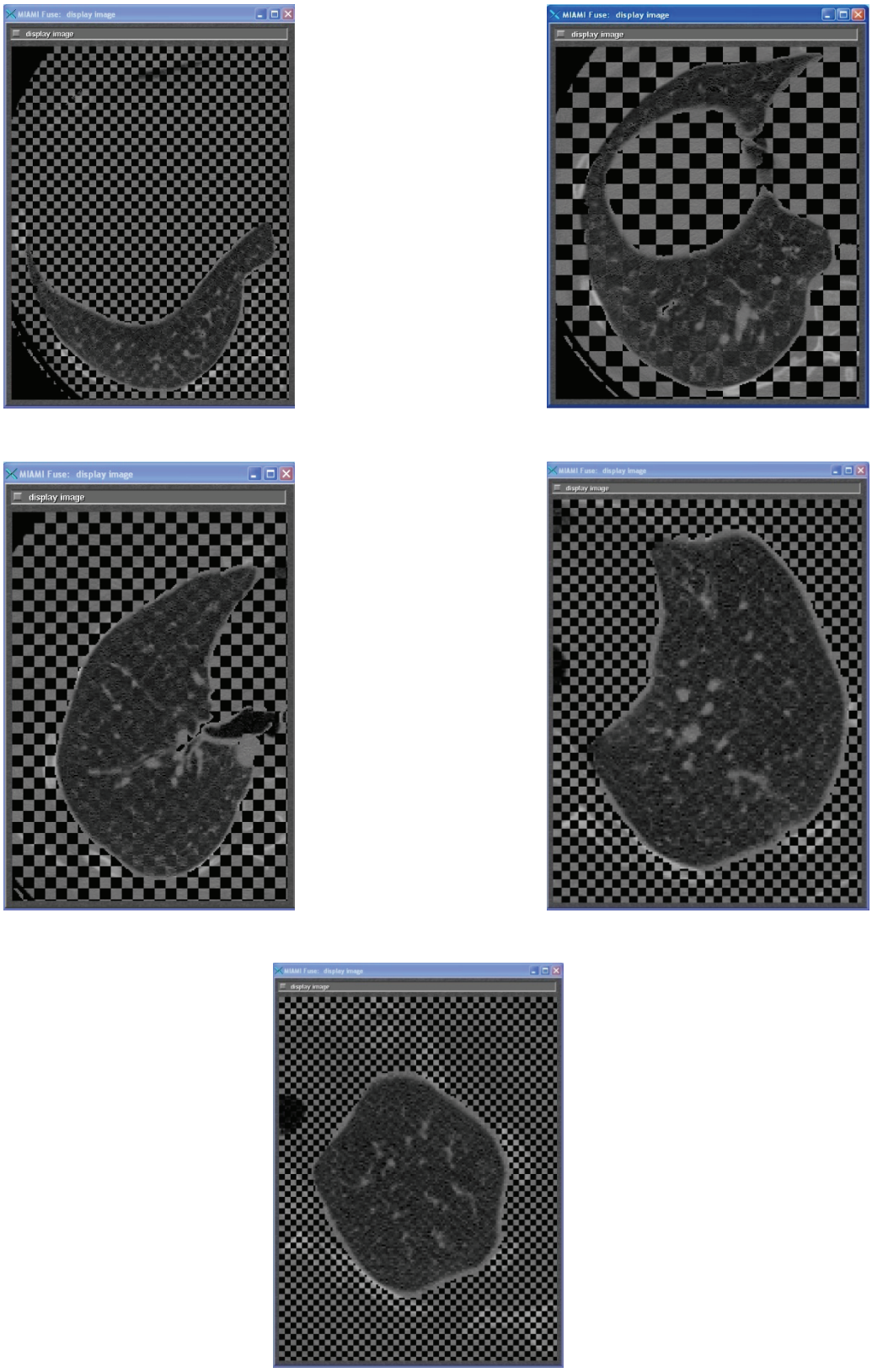
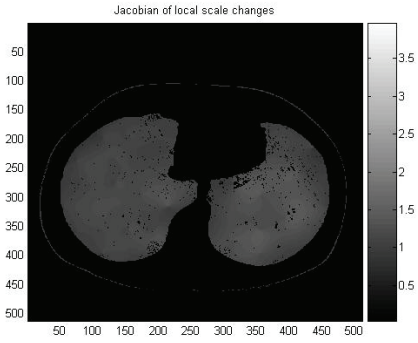
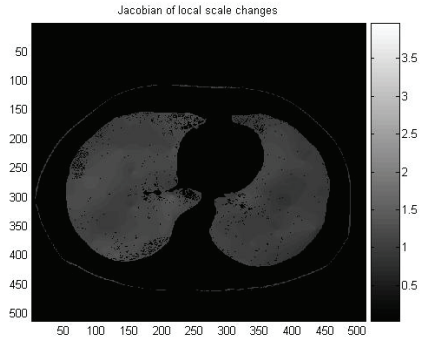


Figure 5-2 Checkerboard plots of registration for “normal” patient

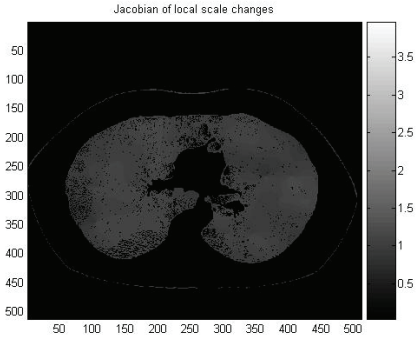
Note that not only is there a good global alignment but the internal features are also well registered. This is desirable if we are to use registration to accurately account for both local and global changes in lung volume. Figure 5-3 and Figure 5-4 show the Jacobian of the computed deformation in representative slices for the Grade 1 and Grade 4 patient respectively. Some deep black spots visible inside the lungs are because features in these regions were removed by the mask obtained from the region growing segmentation algorithm. These regions were not included in the registration and the Jacobian values of deformation there have been set to 0 for convenience. Note that a clear difference in the Jacobian values is visible by comparing Figure 5-3 with Figure 5-4 (means and standard deviations of Jacobian values included). The Jacobian values for the Grade 4 "normal" patient are visibly higher than that of the Grade 1 patient with moderate to severe disease. This is because the local expansions in lung volume are higher for a "normal" patient when going from the expiration phase to the inspiration phase than for a patient with obstructive disease.



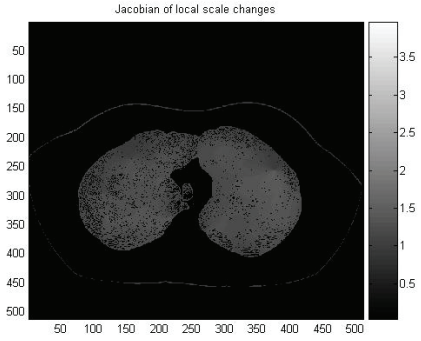
(a) Jacobian values: 1.15 ± 0.11



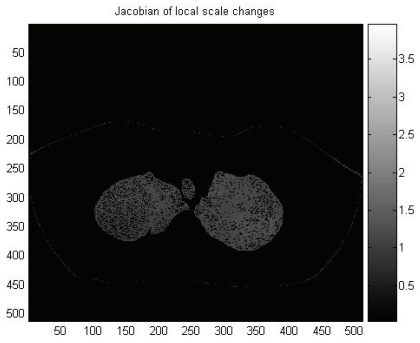
(b) Jacobian values: 1.07 ± 0.09



(c) Jacobian values: 1.02 ± 0.07

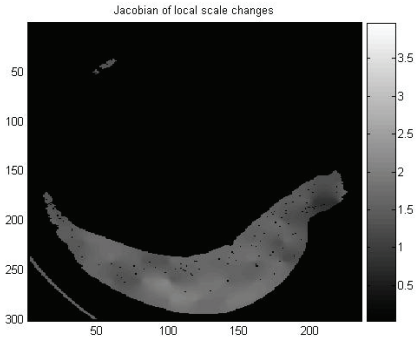


(d) Jacobian values: 1.21 ± 0.11

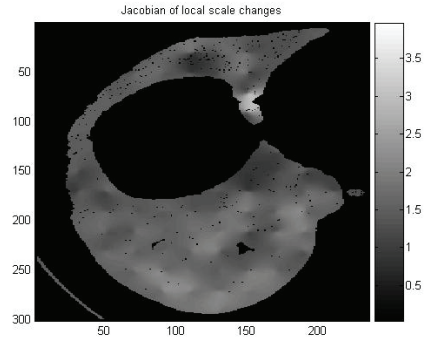


(e) Jacobian values: 1.11 ± 0.07

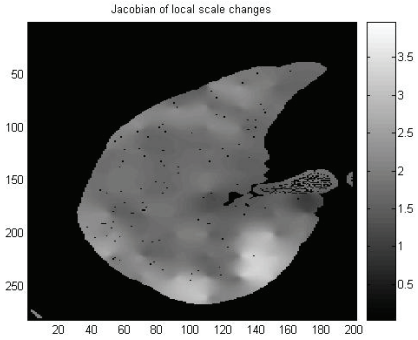
Figure 5-3 Jacobian values of local scale changes in representative slices for Grade 1 patient with moderate to severe disease. Mean and standard deviation of the Jacobian values over the lung mask are included in (a) through (e)



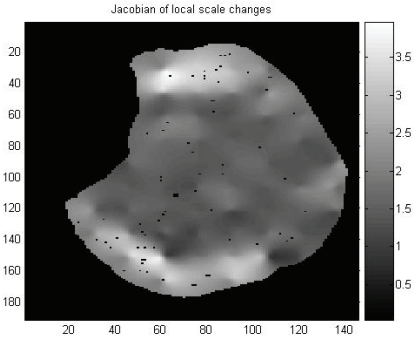
(a) Jacobian values: 1.57 ± 0.21



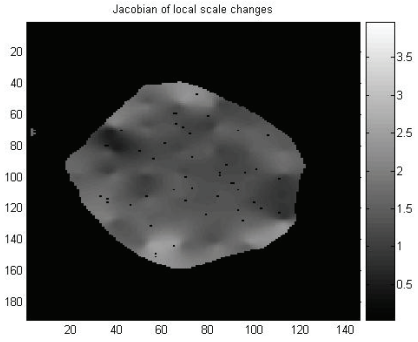
(b) Jacobian values: 1.54 ± 0.28



(c) Jacobian values: 1.90 ± 0.38



(d) Jacobian values: 1.84 ± 0.56



(e) Jacobian values: 1.45 ± 0.35

Figure 5-4 Jacobian values of local scale changes in representative slices for Grade 4 patient classified clinically as "normal". Mean and standard deviation of the Jacobian values over the lung mask are included in (a) through (e)

The figures above indicate that the Jacobian of the deformation of the registration can potentially be used to track chronic obstructive pulmonary disease. The changes in lung volume during the respiratory cycle are correlated with the stage of obstructive disease. An accurate volume change measurement algorithm, such as the one developed and discussed over the last few chapters, should potentially be capable of tracking such changes. Many more patients at varying stages of lung disease would need to be analyzed to see if statistically significant differences can be detected in the Jacobian values of local scale changes for different stages of the disease. If it is possible to statistically detect differences, a lung atlas of Jacobian values can then potentially be used for tracking the progression of the disease in a longitudinal study.

Chapter 6

Conclusion

6.1 Contributions

This thesis advances the efforts made towards improving the sensitivity of tumor size change as an anatomical marker of tumor response. Using RECIST 1.1, the standard technique for response assessment, a spherical tumor's volume needs to change by as much as -65% to be classified as partial response or by as much as 73% before it falls under the progressive disease category [4]. These numbers suggest that while the specificity of RECIST 1.1 is likely high (i.e., tumors that have not responded or have not shown progression are unlikely to have such large volume changes and be incorrectly classified in the partial response or progressive disease category), it is hardly sensitive in terms of *early* detection, i.e., even if a tumor has responded or shown progression it will not be detected unless its volume changes by around 70%. This implies that potentially valuable time and resources are wasted before a therapy regimen can be classified as either effective or ineffective. Low sensitivity also implies a high false negative rate (i.e., the tumor has responded or shown progression but is incorrectly classified as stable) which adversely affects the negative predictive value of RECIST 1.1. Among the

methods being designed for tumor volume change measurement, the one best suited for early change detection would have the least measurement noise and least bias on average in terms of measuring volume change [21]. Most methods for tumor volume change quantification involve the sequential segmentation of the tumor in the two interval scans and then the indirect measurement of change via the subtraction of the two segmented volumes. In this thesis, a low noise, low bias algorithm to measure tumor volume change directly using 3D non-linear image registration has been developed. The registration-based volume change measurement method was tested on synthetic datasets with known non-zero volume change and no statistically significant evidence of functional bias was found in the change measurements. It was also tested in a zero change dual baseline in vivo study of diffusion MRI breast tumors. Two similarity measures were independently used for the registrations – mutual information (MI) and normalized cross correlation (NCC). The 95% confidence interval of error in measured volume change was (-8.93%, 10.49%) with a mean bias of 0.78% for MI-based registration and (-7.69%, 8.83%) with a mean bias of 0.57% for NCC-based registration. Zero was not excluded from the 95% confidence intervals for the mean bias of both MI and NCC-based registrations indicating that our volume change estimation was either unbiased or the tests were insufficiently powered to detect a bias. The 95% confidence intervals stated above for the error in measured volume change are *substantially* more sensitive than RECIST. This implies that if appropriate thresholds for response classification are chosen for the developed registration-based method, the specificity of RECIST 1.1 can potentially be matched at a much higher sensitivity, i.e., a receiver operating characteristic curves analysis would likely have a higher area under the curve for the registration-based method than RECIST

1.1. To the best of this author's knowledge, the error bounds reported in this thesis are also the tightest bounds reported thus far for zero change *in vivo* studies [46], [67], [71], [80]. This implies that registration-based methods may potentially have higher sensitivity than indirect segmentation-based change measurement methods. Additional testing, however, needs to be done on common databases with known truth to compare the performance of the developed algorithm with other methods.

A linear mixed effects model was developed to characterize the performance of the registration-based change measurement algorithm. The model shows that the error in measuring volume change using the registration-based method increases with increase in tumor volume. It also indicates that both volume change error and volume change percentage error decrease with the increase in normalized mutual information content of the tumor even when that is not the similarity measure being optimized during registration.

Approaches that calculate change indirectly by independent segmentation of the tumors in the two exams ignore crucial mutual information present in the tumors. The developed registration-based change measurement algorithm is compared with segmentation-based methods of volume change measurement [66], [71] to demonstrate that the registration-based method is potentially capable of outperforming indirect segmentation-based volume change measurement algorithms. In addition to providing a summary statistic of volume change, a registration-based approach can also potentially show regions of differential growth and contraction across the lesions that are difficult to obtain via binary segmentation approaches.

The proposed registration-based volume change measurement algorithm is not limited to the estimation of volume change in tumors. A potential application is in tracking the progress of obstructive lung disease that affects the changes in lung volume during respiration. The Jacobian of the deformation obtained from the registration can potentially be used to track such changes and relate them to the state of the disease. This possibility is demonstrated in a preliminary feasibility study of a patient diagnosed with chronic obstructive pulmonary disease versus a normal patient.

In summary,

- A low noise, low bias tumor volume change measurement algorithm using 3D nonlinear registration has been developed.
- A linear mixed effects model has been developed to show that using the proposed algorithm the error in measuring volume change increases with increase in the volume of the tumor and decreases with the increase in its normalized mutual information content even when that is not the similarity measure being optimized.
- The proposed registration-based change measurement algorithm has been compared with other approaches and it has the potential to outperform indirect segmentation-based change measurement methods.
- The potential of an accurate registration-based volume change measurement algorithm in tracking progression of chronic obstructive pulmonary disease has been suggested.

6.2 Future Work

6.2.1 General concerns

The goal of this thesis and other similar works is to develop accurate tools for measuring tumor size change. The underlying assumption here is that anatomical size change is indicative of the patient's response to therapy. While this may be true in general for cytotoxic therapy, anatomical markers like tumor size change are alone not enough to categorize the response of tumors to the broad spectrum of therapy drugs [49]. In many cases functional biomarkers like changes in the apparent diffusion coefficient in Diffusion-weighted MRI exams [81] or PET exam studies [12] may be more indicative of response. The need then, in future, is to not only develop tools for accurately measuring anatomical size change but also to investigate if the measured changes have meaning in terms of clinical response.

6.2.2 Tumor volume change measurement using compactly supported RBF

Thin plate splines were used as a deformation interpolant in the registrations described in this thesis. Most registration algorithms start out with an initial step in which the two images are globally registered using affine or rigid body transformation followed by non-linear warping at finer scales. Thin plate splines (TPS) are a globally supported radial basis function with a built-in affine component and hence do not need this additional step during registration. However, being globally supported, computation times for thin plate splines are usually higher. An alternative is to use compactly supported radial basis functions (RBF) like Wu's [28] or Wendland's [29] functions. Besides the gain in computation time, it will be interesting to see how a local RBF does in tracking local

scale changes of tumors across intervals. These local RBFs and the globally supported TPS have been discussed in Chapter 1 section 1.2.1.2 of this thesis.

Note that I attempted to use Wendland's $\psi_{3,2}$ function as a deformation interpolant for registering tumors across interval scans. However, while images that were previously aligned globally were successfully registered, Wendland's function was unable to capture deformations without such a preliminary global alignment. In other words, while the compactly supported RBF did well in capturing local deformation, the registration folded in the cases where there was global misalignment.

Wendland's functions, unlike TPS, do not require the polynomial affine part and the warping coefficients $\{W_i\}$ are solved as

$$KW = F(x_U)$$

where K is a $N \times N$ matrix of radial basis function expansions with

$$K_{ij} = \psi\left(\left\|x_{U\boldsymbol{i}} - x_{U\boldsymbol{j}}\right\|\right)$$

A future plan is to augment the K matrix of the compactly supported RBF to support additional degrees of freedom in the coefficient matrix W so that a global affine solution together with the warping coefficients can be computed. The resulting ill-posed set of equations can be solved using a pseudo-inverse of the K matrix to obtain a least squares solution.

Bibliography

1. Zubrod CG, Scheiderman SM, Frei E III, Brindley C, Gold GL and Schnider B, *Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and thio phosphoamide.* Journal of Chronic Diseases, 1960. **11**: p. 7-33.

2. *WHO handbook for reporting results of cancer treatment*. 1979. Geneva, Switzerland: World Health Organization Offset Publication.

3. Miller AB, Hogestraeten B, Staquet M and Winkler A, *Reporting results of cancer treatment.* Cancer, 1981. **47**: p. 207-14.

4. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Glabbeke MV, van Oosterom AT, Christian MC and Gwyther SG, *New guidelines to evaluate the response to treatment in solid tumors.* Journal of the NCI, 2000. **92**: p. 205-16.

5. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A and Therasse P, *Measuring response in solid tumors: unidimensional versus bidimensional measurement.* J Natl Cancer Inst, 1999. **91**(6): p. 523-528.

6. Trillet-Lenoir V, *Assessment of tumour response to chemotherapy for metastatic colorectal cancer: accuracy of the RECIST criteria.* The British Journal of Radiology, 2002. **75**: p. 903-908.

7. Park JO, Lee S II, Song SY, Kim K, Kim WS, Jung CW, Park YS, Im Y, Kang WK, Lee MH, Lee KS and Park K, *Measuring response in solid tumors: Comparison of RECIST and WHO response criteria.* Jpn J Clin Oncol, 2003. **33**(10): p. 533-537.

8. Jaffe CC, *Response Assessment in Clinical Trials Implications for Sarcoma Clinical Trial Design.* The Oncologist, 2008. **13**(suppl 2): p. 14-18.

9. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D and Verweij J, *New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1).* EJC, 2009. **45**: p. 228-247.

10. Kimura M and Tominaga T, *Outstanding Problems with Response Evaluation Criteria in Solid Tumors (RECIST) in Breast Cancer.* Breast Cancer, 2002. **9**(2): p. 153-159.

11. McHugh K and Kao S, *Response evaluation criteria in solid tumours (RECIST): problems and need for modifications in paediatric oncology?* The British Journal of Radiology, 2003. **76**: p. 433-436.

12. Zhao B, Schwartz LH and Larson SM, *Imaging Surrogates of Tumor Response to Therapy: Anatomic and Functional Biomarkers.* J Nucl Med, 2009. **50**: p. 239-249.

13. Partridge SC, Gibbs JE, Ly Y, Esserman LJ, Tripathy D, Wolverton DS, Rugo HS, Hwang ES, Ewing CA and Hylton NM, *MRI measurements of breast tumor volume predict response to neoadjuvant chemotherapy and recurrence-free survival. AJR*, 2005. **184:**: p. 1774-1781.

14. Yeo SG, Kim DY, Kim TH, Jung KH, Hong YS, Chang HJ, Park JW, Lim S, Choi HS and Jeong S, *Tumor volume reduction rate measured by magnetic resonance volumetry correlated with pathologic tumor response of preoperative chemotherapy for rectal cancer. Int. J. Radiation Oncology Biol. Phys.*, 2010. **78**(1): p. 164-171.

15. Levine ZH, Borchardt BR, Brandenburg NJ, Clark CW, Muralikrishnan B, Shakarji CM, Chen JJ and Siegel EL, *RECIST versus volume measurement in medical CT using ellipsoids of known size.* Optics Express, 2010. **18**(8): p. 8151-8159.

16. Fox J, Ford E, Redmond K, Zhou J, Wong J and Song DY, *Quantification of tumor volume changes during radiotherapy for non-small-cell lung cancer.* Int. J. Radiation Oncology Biol. Phys., 2009. **74**(2): p. 341–348.

17. Tsien C, Gomez-Hassan D, Haken RKT, Tatro D, Junck L, Chenevert TL and Lawrence T, *Evaluating changes in tumor volume using magnetic resonance imaging during the course of radiotherapy treatment of high-grade gliomas: implications for conformal dose-escalation studies.* Int. J. Radiation Oncology Biol. Phys., 2005. **62**(2): p. 328-332.

18. Kirby J and Freymann J. *RIDER*. 2010; Available from: https://wiki.nci.nih.gov/display/CIP/RIDER.

19. Wormanns D, Kohl G, Klotz E, Marheine A, Beyer F, Heindel W and Diederich S, *Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility.* Eur Radiol, 2004. **14**: p. 86-92.

20. Kuhnigk JM, Dicken V, Bornemann L, Bakai A, Wormanns D, Krass S and Peitgen H, *Morphological Segmentation and Partial Volume Analysis for Volumetry of Solid Pulmonary Lesions in Thoracic CT Scans.* IEEE Trans Med Imaging, 2006. **25**(4): p. 417-434.

21. Meyer CR, Armato SG, Fenimore CP, McLennan G, Bidaut LM, Barboriak DP, Gavrielides MA, Jackson EF, McNitt-Gray MF, Kinahan PE, Petrick N and Zhao B, *Quantitative imaging to assess tumor response to therapy: common themes of measurement, truth data, and error sources.* Transl. Oncol., 2009. **2**(4): p. 198-210.

22. Zitova B and Flusser J, *Image registration methods: a survey.* Image Vis. Comput., 2003. **21**: p. 977-1000.

23. Hill DLG, Batchelor PG, Holden M and Hawkes DJ, *Medical image registration.* Physics in Medicine and Biology, 2001. **46**: p. 1-45.

24. Holden M, *A review of geometric transformations for nonrigid body registration.* IEEE Trans Med Imaging, 2008. **27**(1): p. 111-128.

25. Broit C, *Optimal registration of deformed images*, in *Computer and information science*. 1981, University of Pennsylvania.

26. Christensen GE, Rabbitt RD and Miller MI, *3D Brain mapping using a deformable neuroanatomy.* Phys. Medicine Biol., 1994. **39**: p. 609-618.

27. Bookstein FL, *Principal Warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence* 1989. **11**(6): p. 567-585.

28. Wu Z, *Multivariate compactly supported positive definite radial functions.* Advances in Computational Mathematics, 1995. **4**: p. 283-292.

29. Wendland H, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree.* Advances in Computational Mathematics, 1995. **4**: p. 389-396.

30. Fornefett M, Rohr K and Steihl HS, *Radial basis functions with compact support for elastic registration of medical images.* Image Vis. Comput., 2001. **19**: p. 87-96.

31. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO and Hawkes DJ, *Non-rigid registration using free-form deformations: Application to breast MR images.* IEEE Trans Med Imaging, 1999. **18**(8): p. 712-721.

32. Pluim JPW, Maintz JBA and Viergever MA, *Mutual-Information-Based Registration of Medical Images: A Survey.* IEEE Trans Med Imaging, 2003. **22**(8): p. 986-1004.

33. Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P and Marchal G. *Automated multi-modality image registration based on information theory*. in *Information Processing in Medical Imaging*. 1995. Dordrecht, The Netherlands.

34. Tsao J, *Interpolation artifacts in multimodality image registration based on maximization of mutual information.* IEEE Trans Med Imaging, 2003. **22**(7): p. 854-864.

35. Viola PA, *Alignment by maximization of mutual information*. 1995, Massachusetts Inst. Technol., MA: Boston.

36. Meyer CR, Boes JL, Kim B, Bland PH, Zasadny KR, Kison PV, Koral K, Frey KA and Wahl RL, *Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations.* Med. Image. Anal., 1997. **1**: p. 195-206.

37. Studholme C, Hill DLG and Hawkes DJ, *An overlap invariant entropy measure of 3D medical image alignment.* Pattern Recognition, 1999. **32**: p. 71-86.

38. Pluim JPW, Maintz JBA and Viergever MA, *f-Information Measures in Medical Image Registration.* IEEE Trans Med Imaging, 2004. **23**(12): p. 1508-1516.

39. Maes F, Vandermeulen D and Suetens P, *Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information.* Med. Image Anal., 1999. **3**(4): p. 373-386.

40. Bhagalia R, Fessler JA and Kim B, *Gradient based image registration using importance sampling.* ISBI, 2006: p. 446-449.

41. Nelder JA and Mead R, *A simplex method for function minimization.* Computer Journal, 1965. **7**: p. 308–313.

42. Sarkar S, Narayanan R, Park H, Ma B, Bland PH and Meyer CR. *Quantitative growth measurement of lesions in hepatic interval CT exams*. in *Proc. SPIE*. 2008. San Diego.

43. *Surveillance epidemiology and end results*. Available from: http://seer.cancer.gov/.

44. Thirion JP and Calmon G, *Deformation analysis to detect and quantify active lesions in 3D medical image sequences.* IEEE Trans Med Imaging, 1991. **18**: p. 429-441.

45. Thirion JP, *Image matching as a diffusion process: An analogy with maxwell's demons.* Med. Image Anal., 1998. **2**(3): p. 243-260.

46. Reeves AP, Chan A, Yankelevitz DF, Henschke CI, Kressler B and Kostis WJ, *On measuring the change in size of pulmonary nodules.* IEEE Trans Med Imaging, 2006. **25**: p. 435-450.

47. Meyer CR, Park H, Balter JM and Bland PH, *Method for quantifying volumetric lesion change in interval liver ct examinations.* IEEE Trans Med Imaging, 2003. **22**: p. 776-781.

48. DiCiccio TJ and Efron B, *Bootstrap confidence intervals.* Statistical Science, 1996. **11**: p. 189-228.

49. Meyer CR, Park H, Ma B, Kim B and Bland PH, *Potential roles for retrospective registration in molecular imaging*, in *Molecular Imaging: Principles and Practice*, Ralph Weissleder, Editor. p. 262-276.

50. Meyer CR, Johnson TD, McLennan G, Aberle DR, Kazerooni EA, MacMahon H, Mullan BF, Yankelevitz DF, Beek EJRV, Armato III SG, McNitt-Gray MF, Reeves AP, Gur D, Henschke CI, Hoffman EA, Bland PH, Laderach G, Pais R, Qing D, Piker C, Guo J, Starkey A, Max D, Croft BY and Clarke LP, *Evaluation of lung mdct nodule annotation across radiologists and methods.* Academic Radiology, 2006. **13**: p. 1254-1265.

51. Gavrieldes MA, Kinnard LM, Myers KJ, Zeng R and Petrick N. *FDA phantom CT database: a resource for the assessment of lung nodule size estimation methodologies and software development*. in *Proc. of SPIE*. 2010.

52. Clatz O, Sermesant M, Bondiau P, Delingette H, Warfield SK, Malandain G and Ayache N, *Realistic simulation of the 3d growth of brain tumors in mr images coupling diffusion with biomechanical deformation.* IEEE Trans Med Imaging, 2005. **24**(10).

53. Press WH, Flannery BP, Teukolsky SA and Vetterling WT, *Numerical Recipes in C: The Art of Scientific Computing*. 1988: Cambridge University Press.

54. Schwarz GE, *Estimating the dimension of a model.* Annals of Statistics, 1978. **6**(2): p. 461-464.

55. Bland JM and Altman DG, *Measuring agreement in method comparison studies.* Stat. Methods Med. Res., 1999. **8**(135): p. 135-160.

56. Hunter RD, *WHO handbook for reporting results of cancer treatment.* Int. J. Radiat. Biol., 1980. **38**(4): p. 481.

57. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC and Gerig G, *User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability.* Neuroimage, 2006. **31**(3): p. 1116-28.

58. Zhu SC and Yuille A, *Region competition: unifying snakes, region growing, and Bayes/mdl for multiband image segmentation.* IEEE Trans Pattern Anal Mach Intell, 1996. **18**(9): p. 884-900.

59. Sethian JA, *Level set methods and fast marching methods*. 1999: Cambridge Univ. Press.

60. Gavrieldes MA, Kinnard LM, Myers KJ and Petrick N, *Noncalcified Lung Nodules: Volumetric Assessment with Thoracic CT.* Radiology, 2009. **251**(1): p. 26-37.

61. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B and Henschke CI, *Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation.* Radiology, 2000. **217**: p. 251-256.

62. Okada K, Comaniciu D and Krishnan A, *Robust anisotropic gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT.* IEEE Trans Med Imaging, 2005. **24**(3): p. 409-423.

63. Goo JM, Tongdee T, Tongdee R, Yeo K, Hildebolt CF and Bae KT, *Volumetric Measurement of Synthetic Lung Nodules with Multi–Detector Row CT: Effect of Various Image Reconstruction Parameters and Segmentation Thresholds on Measurement Accuracy.* Radiology, 2005. **235**: p. 850-856.

64. Kostis WJ, Reeves AP, Yankelevitz DF and Henschke CI, *Three-Dimensional Segmentation and Growth-Rate Estimation of Small Pulmonary Nodules in Helical CT Images.* IEEE Trans Med Imaging, 2003. **22**(10): p. 1259-1274.

65. Ko JP, Rusinek H, Jacobs EL, Babb JS, Betke M, McGuinness G and Naidichet DP, *Small Pulmonary Nodules: Volume Measurement at Chest CT—Phantom Study.* Radiology, 2003. **228**: p. 864-870.

66. Krishnan K, Ibanez L, Turner WD, Jomier J and Avila RS, *An open-source toolkit for the volumetric measurement of CT lung lesions.* Optics Express, 2010. **18**(14): p. 15256-15266.

67. Gietema HA, Schaefer-Prokop CM, Mali WPTM, Groenewegen G and Prokop M, *Pulmonary Nodules: Interscan Variability of Semiautomated Volume Measurements with Multisection CT— Influence of Inspiration Level, Nodule Size, and Segmentation Performance.* Radiology, 2007. **245**(3): p. 888-894.

68. Jirapatnakul AC. *Semi-automated measurement of pulmonary nodule growth without explicit segmentation*. in *ISBI*. 2009.

69. *VOLCANO '09*. 2009; Available from: http://www.via.cornell.edu/challenge/.

70. *The Second International Workshop on Pulmonary Image Analysis*. 2009; Available from: http://www.lungworkshop.org/2009/index.html.

71. Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, Qin Y, Riely GJ, Kris MG and Schwartz LH, *Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non–Small Cell Lung Cancer.* Radiology, 2009. **252**: p. 263-272.

72. Gavrieldes MA, Kinnard LM, Myers KJ, Peregoy J, Pritchard WF, Zeng R, Esparza J, Karanian J and Petrick N, *A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom.* Optics Express, 2010. **18**(14): p. 15244-15255.

73. Reeves AP, Biancardi AM, Apanasovich TV, Meyer CR, MacMahon H, van Beek EJ, Kazerooni EA, Yankelevitz D, McNitt-Gray MF, McLennan G, Armato SG 3rd, Henschke CI, Aberle DR, Croft BY and al., *The lung image database consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements.* Academic Radiology, 2007. **14**(12): p. 1475-1485.

74. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, Weel CV and Zielinski J, *Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease: GOLD Executive Summary.* Am. J. Respir. Crit. Care Med., 2007. **176**(6): p. 532-555.

75. Christensen GE, Song JH, Lu W, Naqa IE and Low DA, *Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry.* Med. Phys., 2007. **34**(6): p. 2155-2163.

76. Christensen GE and Johnson H, *Consistent image registration.* IEEE Trans Med Imaging, 2001. **20**(7): p. 568-582.

77. Reinhardt JM, Ding K, Cao K, Christensen GE, Hoffman EA and Bodas SV, *Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation.* Med. Image Anal., 2008. **12**: p. 752-763.

78. Li B, Christensen GE, Hoffman EA, McLennan G and Reinhardt JM, *Pulmonary CT image registration and warping for tracking tissue deformation during the respiratory cycle through 3D consistent image registration.* Med. Phys., 2008. **35**(12): p. 5575-5583.

79. *itk Insight Toolkit*. Available from: http://www.itk.org/.

80. Mozley PD, Schwartz LH, Bendtsen C, Zhao B, Petrick N and Buckler AJ, *Change in lung tumor volume as a biomarker of treatment response: a critical review of the evidence.* Annals of Oncology, 2010. **21**: p. 1751-1755.

81. Ma B, Meyer CR, Pickles MD, Chenevert TL, Bland PH, Galbán CJ, Rehemtulla A, Turnbull LW and Ross BD, *Voxel-by-Voxel Functional Diffusion Mapping for Early Evaluation of Breast Cancer Treatment* Proceedings of Information Processing in Medical Imaging, Lecture Notes in Computer Science (LNCS), 2009. **5636**: p. 276-287.