REGRESSION LEARNING FOR 2D/3D IMAGE REGISTRATION

Chen-Rui Chou

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2013

Approved by:

Stephen M. Pizer

Sha X. Chang

David S. Lalush

J. S. Marron

Marc Niethammer

ABSTRACT

CHEN-RUI CHOU: REGRESSION LEARNING FOR 2D/3D IMAGE REGISTRATION.

(Under the direction of Stephen M. Pizer.)

Image registration is a common technique in medical image analysis. The goal of image registration is to discover the underlying geometric transformation of target objects or regions appearing in two images. This dissertation investigates image registration methods for lung Image-Guided Radiation Therapy (IGRT). The goal of lung IGRT is to lay the radiation beam on the ever-changing tumor centroid but avoid organs at risk under the patient's continuous respiratory motion during the therapeutic procedure.

To achieve this goal, I developed regression learning methods that compute the patients 3D deformation between a treatment-time acquired x-ray image and a treatment-planning CT image (2D/3D image registration) in real-time. The real-time computation involves learning x-ray to 3D deformation regressions from a simulated patient-specific training set that captures credible deformation variations obtained from the patients Respiratory-Correlated CT (RCCT) images. At treatment time, the learned regressions can be applied efficiently to the acquired x-ray image to yield an estimation of the patients 3D deformation. In this dissertation, three regression learning methods - linear, non-linear, and locally-linear regression learning methods are presented to approach this 2D/3D image registration problem.

In honor of
my parents & Ya-Shen

ACKNOWLEDGEMENTS

I'd like to express my sincere appreciation to all of the people who have supported my research and this dissertation. I begin by thanking my advisor, Steve Pizer, for his patience, guidance and supports over this four years. He has set me an example of excellence as a great researcher, mentor, and instructor. I would like to thank all other members of my dissertation committee, Sha Chang, David Lalush, Steve Marron, and Marc Niethammer, and my collaborators, Brandon Frederick and Gig Mageras, for their extremely insightful suggestions on this research. I would like to thank department of Computer Science for giving me admission and financial support for my graduate study. I would also like to thank department of Radiation Oncology for supporting my research and providing me a chance for collaboration. I would like to thank fellows from Image Lunch and Shape Stats for our weekly stimulating discussion and vision-broadening presentations. They have helped me greatly in developing my research ideas. I would like to thank my supervisors during my two internships: Christophe Chefd'Hotel in Siemens Corporate Research and Ben Glocker in Microsoft Research Cambridge for keeping me in doing medical image research during the summer. I would like to thank my three-year officemate, Liang Shan, for our daily discussions, sometimes disputes, over our research ideas. I would also like to thank all my Taiwanese friends in North Carolina, for giving me the best work-life balance that I ever had. Lastly, I'd like to thank my family and my lovely wife Ya-Shen for bearing my five financially-unproductive years and for loving me with no return.

TABLE OF CONTENTS

Chapter 1: Introduction

1.1 Image Registration for Medical Image Analysis

Image registration is a common technique in medical image analysis. The goal of image registration is to discover the underlying geometric transformation of target objects or regions appearing in two images. In many medical situations doctors need to understand transformation of the target (e.g., organs or tissues) between two images before making any medical decisions. Therefore, image registration is a fundamental task of medical image analysis.

Many current medical imaging techniques compute 3D image data by reconstruction of a collection of 2D projection images acquired at various angles, e.g., CT (Computed Tomography). Their 3D image quality (volumetric information) increases with the number of angular samples of projection images. Image registration between two high-quality 3D images has shown good estimation of the underlying transformation (Vercauteren *et al.* (2009); Rohr *et al.* (2001); Christensen *et al.* (1996a); Rueckert *et al.* (1999); Pluim *et al.* (2003)).

However, in medical applications where the target anatomy or object of interest is dynamic, the acquisition and reconstruction of a high-quality 3D image is too slow to represent the target position and shape in real time. In addition, acquiring high-quality 3D CT images throughout a therapeutic procedure may impose extra radiation dose on patients and introduce radiation disease.

1.2 Challenges of Image Registration as Guidance for Radiation Therapy

One medical application of registration is lung Image-Guided Radiation Therapy (IGRT). The goal of lung IGRT is to lay the radiation beam on the ever-changing tumor centroid under the patient's continuous respiratory motion but to avoid organs at risk (OAR) during the therapeutic procedure. Therefore, image registration between high-quality 3D images can not provide timely information of the tumor location.

Recent advances of IGRT registration methods suggest a new direction: 2D/3D image registration. 2D/3D image registration for IGRT estimates the underlying 3D transformation between a high-quality 3D image acquired at planning time and a small set of projections that can be quickly acquired at treatment time.

2D/3D image registration computes 3D transformations based on information of the 2D projection intensities. It estimates a 3D transformation where the projection intensities of the transformed 3D volume match the target projection intensities. Due to the mismatch in the registration dimensions, the 2D/3D image registration problem is ill-posed: the unknowns (transformations at all 3D pixels) are orders of magnitude more than the given constraints (intensity differences at all projection pixels). Therefore, in order to solve the 2D/3D image registration, one needs to impose more spatial or temporal constraints on the transformation.

For rigid transformations, one can easily impose the global transformation constraint and reduce the unknown to the six Euler transformation parameters. For non-rigid transformations, current 2D/3D image registration methods (Li *et al.* (2011a, 2010); Liu *et al.* (2010)) introduce a spatial constraint on the 3D transformations: instead of computing over every possible transformation, the methods compute over a parameterized subspace of the possible transformations. The possible transformation subspace can be formed by linear combinations of the transformation bases observed in the patient's 4D CTs at treatment-planning time (Ch. 4).

At treatment time, with the constrained and parameterized transformations the 2D/3D image registration method now is a more well-posed problem with a low-dimensional unknown space of parameters: optimizing the transformation parameters (dimension < 10) by matching the target projection intensities (dimension $> 100,000$) and the projection intensities of the estimated 3D volume.

Another challenge for registrations in IGRT is the computation time. To have a *real-time*[1] registration method that can deal with more than 10 registrations per second, the optimization-based 2D/3D image registration approach (Li *et al.* (2011a, 2010)) above requires GPU acceleration on the iterative calculation of the objective function's Jacobian and a good initialization of the parameters (Ch. 2). Given the fact that a good initialization of the optimizing parameters is not normally available, in this dissertation I seek a new solution by regression learning that can provide efficient 2D/3D image registration in IGRT without requirements on initialization.

1.3 A Brief Outline of the Proposed Methods

The idea of regression learning is as follows: if we have a way to sample the patient's credible 3D transformations and simulate the corresponding projection intensity patterns, at treatment-planning time we can create a patient-specific training set of projection intensity and transformation pairs and thereby learn a set of regression functions that map projection intensities to 3D transformations. Using this at treatment time, given a projection intensity pattern we can apply those learned regression functions and efficiently estimate the patient's deformation.

In this dissertation, three types of regression functions (linear, nonlinear, and locally-linear regressions) have been investigated for learning from the patient-specific training

---

[1] 10 registrations per second is considered as real-time computation given the fact that current CBCT projection acquisition rate is about 10 Hz Li *et al.* (2011a).

sets.

1.3.1 Linear Regression Learning

The linear-regression-learning method is called Correction via Limited-Angle Residues in External beam Therapy (CLARET), which learns a global linear regression function that maps projection intensities into the transformation parameters. At treatment time the patient's transformation parameters can be estimated iteratively (CLARET-itr) or non-iteratively (CLARET-nonitr) by the projection residues between the target projection and the calculated projection of currently-estimated image (Ch. 5). The purpose of iterating the estimation is to have a more accurate registration. The method has appeared in Chou *et al.* (2010b,a, 2011b,c,a, 2013).

1.3.2 Non-linear Regression Learning

The nonlinear-regression-learning method is called Registration Efficiency and Accuracy through Learning Metrics on Shape (REALMS). To relax the strong linear assumption (between transformation parameters and projection intensities) made in CLARET, REALMS learns distance metrics for non-linear kernel regressions between the transformation parameters and the projection intensities. This method is a non-iterative method: at treatment time, the transformation parameters are interpolated once from the training parameters weighted by kernels equipped with the learned distance metric (Ch. 6). The method has been appeared in Chou *et al.* (2012); Chou and Pizer (2012).

1.3.3 Locally-linear Regression Learning

In addition to the linear and nonlinear regression learning methods, I also investigate the feasibility of learning linear regressions in a local transformation neighborhood, which I call "local-learning", or "locally-linear learning" for the rest of the dissertation.

The local-learning version of the non-iterative CLARET is called Local-CLARET (L-CLARET). The goal of L-CLARET is to find a balance between the strong linear assumption in CLARET and the time-consuming non-linear learning in REALMS. Instead of learning a global linear mapping, L-CLARET learns linear regression mappings for every local training parameter neighborhood to obtain better regression fitting. At treatment time, the acquired projection is first classified into a training local neighborhood by efficient decision forest classification based on projection image visual features. Second, the transformation parameters are estimated with the linear regression of this local training neighborhood (Ch. 7). The method has appeared in Chou and Pizer (2013).

1.4 Thesis and Contributions

Thesis: *Regression learning provides a new solution to the image registration problem. Learning patient-specific intensity-to-shape regressions allows efficient, accurate, and robust 2D/3D image registration for image-guided radiation therapy.*

The contributions of this dissertation are the following:

(1) The development of four regression-learning-based 2D/3D image registration methods for image-guided radiation therapy.

  a. CLARET (Correction via Limited-Angle Residues in External Beam Therapy)

  b. REALMS (Registration Efficiency and Accuracy through Learning Metric on Shape)

  c. L-CLARET (Local CLARET)

(2) The implementation of CLARET, L-CLARET, REALMS and L-REALMS into software.

(3) The iterative versions of the four methods in **(1)**: enhancing registration accuracy by iterative estimation.

(4) The development of scattering removal and intensity correction on the digitally reconstructed radiographs (DRR) and the treatment-time radiographs to allow the commensurate intensity comparison required by the proposed methods.

(5) The evaluation of CLARET, L-CLARET and REALMS for lung IGRT with simulated and real patient cone-beam radiographs, including comparisons to an optimization-based 2D/3D registration approach.

(6) The evaluation of CLARET for head-and-neck IGRT with simulated and real patient Nanotube Stationary Tomosynthesis (NST) (Maltz *et al.* (2009)) radiographs.

1.5 Overview of Chapters

The rest of this dissertation is organized as follows: Chapter 2 describes existing image registration methods. Chapter 3 describes the IGRT clinical background. Chapter 4 describes parameterizing the transformation for IGRTs. Chapters 5, 6 and 7 respectively describe and evaluate the CLARET, the REALMS, and the L-CLARET methods. Chapter 8 describes the comparative evaluations of those methods in the lung IGRT. Chapter 9 summarizes the methods, discusses the results and their problems, and presents future work.

Chapter 2: Image Registration

Image registration, i.e., finding the underlying geometric transformation between two images, is widely used in many fields. That is, given a source image $I_0 : \Omega \subset \mathbb{R}^{\mathrm{d}} \to \mathbb{R}$ (d is the image dimensionality, e.g., $d = 2, 3$) and a target image $I_1 : \Omega \to \mathbb{R}$, find a reasonable transformation map $\phi : \Omega \to \mathbb{R}^{\mathrm{d}}$ such that the transformed source image $I_0(\phi)$ is similar to the target image I_1 (Modersitzki (2004)). In computer vision, for example, scientists use image registration to understand pixel correspondence in stereographic projections for 3D scene reconstruction (Blais and Levine (1995)). In medical image analysis, scientists use image registration to understand possible transformations appearing in a sequence of images for pathological staging of a disease (Fox *et al.* (2001)), population analysis (Lorenzen *et al.* (2006); Cootes *et al.* (2004); Bhatia *et al.* (2004); Davatzikos *et al.* (1996)), atlas based segmentations (Isgum *et al.* (2009); Aljabar *et al.* (2009); Collins and Evans (1997); Wu *et al.* (2007)), or aligning images from multiple modalities of the same patient (D'Agostino *et al.* (2003); Gaens *et al.* (1998); Roche *et al.* (1998); Maes *et al.* (1997)). With its popularity, there is a variety of image registration methods developed for various purposes. The classic variational approach for image registration can be formulated as an energy minimization process that finds a displacement field $u : \Omega \to \mathbb{R}^{\mathrm{d}}$ between two images through minimizing an energy function consisting of a data attachment term E_D and a regularization term E_R (Hajnal and Hill (2010); Modersitzki (2004); Broit (1981); Bajcsy and Broit (1982)). The following is an example of an image registration framework:

$$u = \underset{u}{arg\min}\; E_D(I_0(\phi),\; I_1) + \lambda E_R(u), \tag{2.0.1}$$

where $\phi = \mathbf{Id} + u$; λ is the Lagrange multiplier (Bellman (1986)). The data attachment term measures the image dissimilarity using some distance measure on intensities: L_2-norm (Belongie *et al.* (2002); Rueckert *et al.* (1999)), cross-correlation (Lewis (1995); Roche *et al.* (1998)), mutual information (Pluim *et al.* (2003); Maes *et al.* (1997); Gaens *et al.* (1998); D'Agostino *et al.* (2003)). The regularization term uses a physical model that regularizes either the output displacement field: elastic model (Rohr *et al.* (2001)) and diffusion model (Horn and Schunck (1981)) or the time-varying velocity field: fluid model (Christensen *et al.* (1996a)). With this framework, image registration can be classified into two types: methods using a parameterized transformation (2.1) and methods using non-parameterized transformation (2.2).

2.1 Parametric Registration

Image registration without regularization, i.e., led solely by the data attachment energy E_D in Eq. 2.0.1, is ill-posed: the dimension of the unknown displacements is d-fold more than the dimension of the known intensity pairs (source and target). To reduce the unknown dimension and make the registration problem more well-posed (#unknowns$\leq$#knowns), one can parameterize the transformation to obtain a low-dimensional descriptor of the displacement. Thus, instead of optimizing over the entire displacement field, the registration now only requires optimization over that limited number of parameters. Two basic parametric registration methods are rigid and affine registrations. Transformations in rigid (Maes *et al.* (1997)) and affine (Jenkinson and Smith (2001)) registrations can be modeled globally by 6 Euler parameters and 12 affine parameters, respectively ($d = 3$). Most parametric registration methods model transformations by linear models (B-Splines: Rueckert *et al.* (1999, 2006), Principal

Component Analysis (PCA): Li *et al.* (2011a); Liu *et al.* (2010); Chou *et al.* (2013), trigonometric functions: Reddy and Chatterji (1996); Chen *et al.* (1994), Radial Basis Functions (RBF): Chui and Rangarajan (2003); Fornefett *et al.* (2001); Bookstein (1989)). In the following sections, I respectively describe the global parameterization by PCA (Section 2.1.1) and the local parameterization by B-Splines (Section 2.1.2).

2.1.1 Global Parameterization by Linear Analysis (PCA)

If there exists a set of displacement fields of the target object/region, one can parameterize the transformation globally by doing linear analysis based on this a prior displacement fields, e.g., PCA. This global parametric method assumes that the displacement field u at pixel/voxel location $\mathbf{x}$ can be represented by m parameters $\mathbf{c} = (c_1, c_2, ..., c_m)$ as the scores on their eigenmode basis functions $\mathbf{b} = \{b_1, b_2, ..., b_m\}$:

$$u(\mathbf{x}, \mathbf{c}, \mathbf{b}) = \sum_{k=1}^{m} c_k b_k(\mathbf{x}), \tag{2.1.1}$$

where $c_k \in \mathbb{R}$ and $b_k : \Omega \to \mathbb{R}^{\mathrm{d}}$.

With this parameterization, we can re-write the optimization in Eq. 2.0.1 as follows:

$$\mathbf{c} = \underset{\mathbf{c}}{arg\min}\ E_D(I_0(\mathbf{Id} + u(\mathbf{c}, \mathbf{b})),\ I_1) + \lambda E_R(u(\mathbf{c}, \mathbf{b})), \tag{2.1.2}$$

If the set of a priori displacement fields was computed via a variational framework with regularization that guarantees smoothness (Liu *et al.* (2010); Chou *et al.* (2013)), then the analyzed eigenmode basis functions and their linear combinations will be smooth too. Therefore, we can drop the regularization term and let the registration be driven solely by data attachment. Data attachment can be measured as the sum of squared differences in intensity for every image voxel/pixel $\mathbf{x}$ (Belongie *et al.* (2002); Rueckert *et al.* (1999)):

$$\mathbf{c} = \underset{\mathbf{c}}{arg\min} \int_{\Omega} (I_0(\mathbf{x} + u(\mathbf{x}, \mathbf{c}, \mathbf{b})) - I_1(\mathbf{x}))^2 d\mathbf{x} \tag{2.1.3}$$

The final displacement field can be computed by Eq. 2.1.1, using the basis functions **b** and the optimized parameters **c**.

The advantage of this parametric method is that the space of transformations can be greatly reduced by PCA (Li *et al.* (2011b); Liu *et al.* (2010); Li *et al.* (2011a)), so the registration optimization problem is well-posed (#unknown transformation parameters $\ll$ #known intensity pairs). With only a few parameters to optimize, the method is also efficient.

The downside of this parametric approach is that it cannot produce a displacement field outside of the space spanned by those eigenmode basis functions. This constraint discourages registration of objects of high shape variability where no prior displacement sets can include all its variations.

2.1.2 Local Parameterization by B-Splines

If there is no a priori knowledge of the object's transformation, one can still parameterize the transformation space by B-Splines (De Boor *et al.* (1978)) and optimize over the B-Spline parameters for registration. For example, Free-Form Deformation (FFD) registration (Rueckert *et al.* (1999, 2006)) uses cubic B-Spline functions to deform an object by manipulating an underlying mesh of control points. The resulting deformation produces a smooth and continuous transformation. FFD registration samples an $n_x \times n_y \times n_z$ mesh of control points $\mathbf{c} = \{c_{1,1,1},\ ...,\ c_{i,j,k},\ ...,\ c_{n_x,n_y,n_Z}\}$ at grid points ($d = 3$) and takes transformations at control points $\phi = \{\phi_{1,1,1},\ ...,\ \phi_{i,j,k},\ ...,\ \phi_{n_x,n_y,n_Z}\}$ (where $\phi:\ \Omega \rightarrow \mathbb{R}^{\mathrm{d}}$) as the parameters. FFD registration minimizes the image dissimilarity by optimizing over the transformation at those control points. Transformations at voxels (x, y, z) other than the control points are interpolated by the four cubic B-Spline functions $\mathbf{b} = \{b_0, b_1, b_2, b_3\}$ (where $b_k : \Omega \to \mathbb{R}$) among the nearby $3 \times 3 \times 3 = 27$ control points.

$$\phi(x,y,z) = \sum_{p=0}^{3}\sum_{q=0}^{3}\sum_{r=0}^{3} b_p(\frac{x}{n_x} - \left\lfloor \frac{x}{n_x} \right\rfloor) b_q(\frac{y}{n_y} - \left\lfloor \frac{y}{n_y} \right\rfloor) b_r(\frac{z}{n_z} - \left\lfloor \frac{z}{n_z} \right\rfloor) \phi_{i+p,\, j+q,\, k+r}, \quad (2.1.4)$$

where $i = \left\lfloor \frac{x}{n_x} \right\rfloor - 1$, $j = \left\lfloor \frac{y}{n_y} \right\rfloor - 1$, $k = \left\lfloor \frac{z}{n_z} \right\rfloor - 1$, $b_0(\Delta) = \frac{(1-\Delta)^3}{6}$, $b_1(\Delta) = \frac{(3\Delta^3 - 6\Delta^2 + 4)}{6}$, $b_2(\Delta) = \frac{(-3\Delta^3 + 3\Delta^2 + 3\Delta + 1)}{6}$, and $b_3(\Delta) = \frac{\Delta^3}{6}$.

The advantages of B-Spline-based FFD registration are (1) it can parameterize the transformation without a priori information, (2) B-Splines are locally controlled, which makes it computationally efficient (that is, it doesn't need to update the whole image volume for a parameter update), and (3) it can optimize over various transformation scales by multi-level registrations using sparse-to-dense control points (Rueckert *et al.* (1999, 2006)). The registration solution is also unique when the number of control points is less than one third of the number of image voxels.

2.2 Non-parametric Registration

In FFD registration, when sampling control points at every image voxel, the resulting transformation will have the most flexibility. Like FFD registration that uses fully-sampled control points, non-parametric registration grants an independent transformation for each voxel. However, this high flexibility makes the registration problem ill-posed. To make it more well-posed, scientists have come up with various regularization approaches that impose physical models to the transformations. Based on the regularization types, they can be briefly categorized into the following registrations.

Diffusion-based registrations (Thirion (1998); Horn and Schunck (1981); Fischer and Modersitzki (2002)) penalize gradient magnitudes of the displacement field in the regularization energy:

$$E_R^{diff}(u) = \frac{1}{2} \sum_{d=1}^{n} \int_\Omega \|\nabla u_d(\mathbf{x})\|^2 \, d\mathbf{x} \tag{2.2.1}$$

Elastic registrations (Rohr *et al.* (2001); Broit (1981); Bajcsy and Kovacic (1989); Christensen *et al.* (1994a,b); Gee *et al.* (1997)) penalize the elastic potential of the displacement field u in the regularization energy:

$$E_R^{elas}(u) = \int_\Omega \frac{\lambda + \mu}{2} \|\nabla \cdot u(\mathbf{x})\|^2 + \frac{\mu}{2} \sum_{d=1}^{n} \|\nabla u_d(\mathbf{x})\|^2 \, d\mathbf{x} \tag{2.2.2}$$

where the constants λ, $u > 0$ describe material properties.

However, both of the above regularization energies discourage transformations of large displacements. The recently popular Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework using the fluid-flow analogy (Christensen *et al.* (1996a); D'Agostino *et al.* (2003); Bro-Nielsen and Cotin (1996); Christensen *et al.* (1996b); Christensen (1994); Dupuis *et al.* (1998)) allows large displacements. Instead of regularizing the displacement fields, it regularizes the time-dependent velocity fields from fluid mechanics:

$$E_R^{fluid}(u) = \int_\Omega \int_{t=0}^{1} \|v_t(\mathbf{x})\|_V \, d\mathbf{x}, \tag{2.2.3}$$

where the flows v_t: $\Omega \to V$ ($t \in [0, 1]$) are time-dependent velocity fields that are elements of a Hilbert space V: $\Omega \to \mathbb{R}^{\mathrm{d}}$ with inner product $< \cdot, \cdot >_V$. The norm $\|v_t\|_V$ can be expressed as $< Lv_t, Lv_t >_{L_2}$ regularized by the differential operator L frequently taken from fluid mechanics: $L = \alpha\nabla^2 + \beta(\nabla\cdot)\nabla + \gamma$ and $\alpha, \beta, \gamma > 0$. The α term controls the smoothness of the transformation, the β term controls the compressibility of the fluid, and γ penalizes the total size of the transformation.

The fluid-flow-based LDDMM framework can be written in the following form (Beg *et al.* (2005)):

$$v = \underset{v\,:\,\dot{\phi}_t = v_t(\phi_t)}{arg\ \ min} \int_\Omega \int_{t=0}^{1} \|v_t(\mathbf{x})\|_V \, dt + \frac{1}{\sigma^2} \left\| I_0(\phi_{t=1}^{-1}(\mathbf{x})) - I_1(\mathbf{x}) \right\|_{L_2} d\mathbf{x} \qquad (2.2.4)$$

where $\phi_{t=1}(\mathbf{x})$ describes the final transformation ($t = 1$) at the voxel location $\mathbf{x}$ in the "target" image I_1. The transformation at time $\mathbf{T}$ of a voxel location $\mathbf{x}$ can be integrated as follows:

$$\phi_{\mathbf{T}}(\mathbf{x}) = \phi_{t=0}(\mathbf{x}) + \int_0^{\mathbf{T}} v_t(\phi_t(\mathbf{x}))dt \qquad (2.2.5)$$

where $\phi_{t=0}(\mathbf{x}) = \mathbf{x}$.

"As shown in Dupuis *et al.* (1998) and Trouve (1995), enforcing a sufficient amount of smoothness on the elements of the space V of allowable velocity vector fields ensures that the solution to the differential equation $\dot{\phi}_t = v_t(\phi_t)$, $t \in [0, 1]$, $v_t \in V$ is in the space of diffeomorphisms."[1] Therefore, the solution satisfying Eq. 2.2.4 is an LDDMM solution in the following two senses: (1) as shown in Christensen *et al.* (1996b) the fluid-flow approach provides a large deformation coordinate system transformation, and (2) as shown in Miller and Younes (2001), Trouve (1995) and Miller *et al.* (2002), in contrast to Christensen *et al.* (1996b), the length of the shortest path $\inf \int_{t=0}^{1} \|v_t(\mathbf{x})\|_V \, dt$ connecting images I_0 to I_1 defines a metric in the image space.

However, LDDMM is not the only method that produces diffeomorphic transformations. Many other registration methods (Vercauteren *et al.* (2009); Rueckert *et al.*

---

[1] Quoted from Beg *et al.* (2005).

(2006); Ashburner (2007)) have been recently revised to guarantee large and diffeomorphic transformations by regularizing the time-dependent velocity fields.

For medical image analysis, it is of considerable interest to compute transformations that preserve properties such as smoothness of curves, surfaces or other features associated to anatomy. Therefore, diffeomorphic transformations, which are smooth transformations with smooth inverse, are of considerable interest as curves, surfaces, and anatomy topology is preserved, and coordinates are transformed consistently. In Ch. 4 I use this LDDMM framework to discover diffeomorphic transformations among a time series of 3D images.

Chapter 3: Image-Guided Radiation Therapy

3.1 Treatment planning[1]

A CT image taken of patients in the treatment position is used for the radiotherapy treatment planning process. Using this image, the tumor and the normal organ structures at risk are segmented, the radiation field is designed, and the radiation dose is computed. This reference CT should contain information as to where volumetrically the tissue to be treated is. There are three main volumes in radiotherapy planning. The first is the position and the extent of gross tumor, i.e., what can be seen or imaged; this is known as the gross tumor volume (GTV). Developments in imaging have contributed to the definition of the GTV. The second volume contains the GTV, plus a margin for sub-clinical disease spread which therefore cannot be fully imaged; this is known as the clinical target volume (CTV). The CTV is important because this volume must be adequately treated to achieve cure. The third volume, the planning target volume (PTV), allows for uncertainties in planning or treatment delivery. It is a geometric concept designed to ensure that the radiotherapy dose is actually delivered to the CTV. See Fig. 3.1.1 for visualization of those target volumes. Radiotherapy planning must also consider critical normal tissue structures, known as organs at risk (OAR). In some specific circumstances, it is necessary to add a margin analogous to the PTV margin around an OAR to ensure that the organ cannot receive a higher-than-safe

---

[1] The major component of this section is quoted from Burnet *et al.* (2004)

dose; this gives a planning organ at risk volume. This applies to an organ such as the spinal cord, where damage to a small amount of normal tissue would produce a severe clinical manifestation (Burnet *et al.* (2004)).

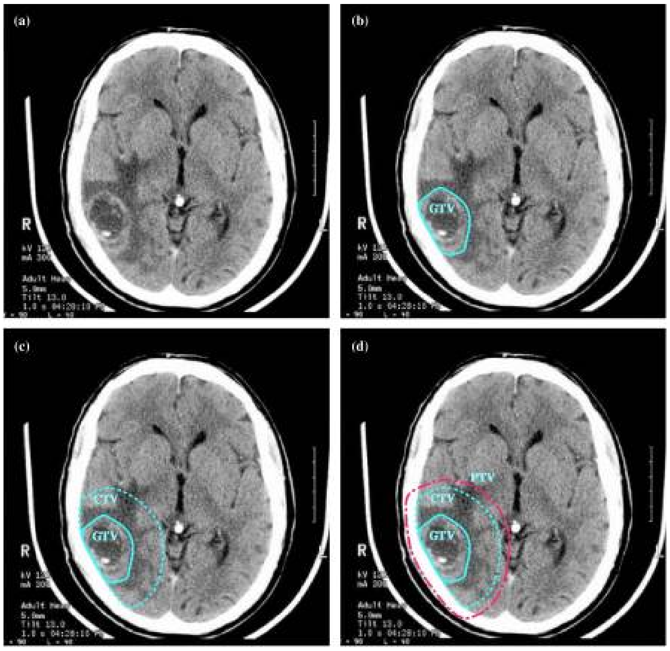

Figure 3.1.1: Planning volumes for a patient with a WHO grade 4 glioma (glioblastoma). (a) Planning CT showing a contrast-enhancing tumor. (b) The GTV is the visible tumor. (c) A margin for microscopic spread has been added to make the CTV; the margin is the same in all directions except that it is restricted by the skull. (d) The PTV has added to the CTV to account for uncertainties in planning and execution of treatment; this extends beyond the inner table of the skull. (Burnet *et al.* (2004))

After delineating the target volumes and the organs at risk volumes, the trajectories and modulation (cross-sectional radiation pattern) of the treatment radiation beam can be inversely planned. That is, given a set of PTVs and OARs, the plan must specify several treatment trajectories and modulation from various angles such that the accumulated radiation dose at those PTVs is at the prescribed level and the accumulated radiation dose at those OARs is at a safe level. See Fig. 3.1.2 for illustration.

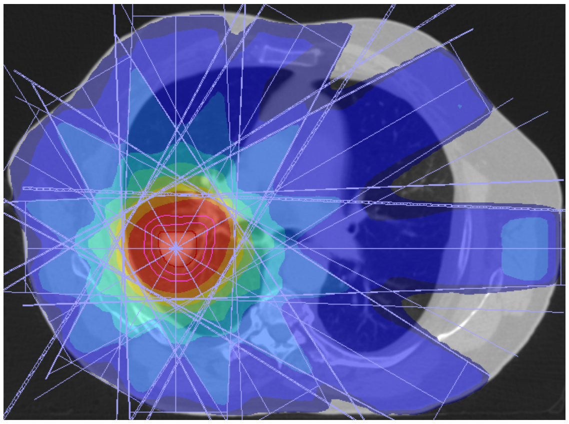

Figure 3.1.2: Treatment planning through simulation of the radiation. The outermost pink contour is the PTV. (Vandemeulebroucke *et al.* (2009))

3.2 Treatment-time Image Guidance

Fig. 3.2.1 shows a typical IGRT environment: an adjustable patient couch and a rotating Cone-Beam CT imager (CBCT, 3.3.1) mounted on a linear accelerator. Before radiation treatment, a 3D image acquired by the CBCT imager will be used to measure the patient's setup deviation between the planning and treatment CT volumes. The amount of setup deviation can be calculated by 3D/3D image registration of the volume of interest to allow accurate setup for daily treatment. The setup deviation can then be corrected rigidly by shifting and rotating the couch position and orientation such that the CT volumes match.

The patient's intra-fraction movements or deformations can be detected by the CBCT imager as well. Radiation oncologists can ask the patient to hold his/her breath, take sequential x-ray projections for 3D image reconstruction, and determine if the target volume is inside the planned region. The other trending approach takes advantage of (1) fast x-ray projection acquisition and (2) fast image registration by GPU acceleration and estimates the patient's transformation directly from x-ray projections in real-time.

Figure 3.2.1: An Elekta radiation treatment machine (Capital Radiation Therapy (2013)) with an x-ray imager mounted.

The clinical goal for registration accuracy is 2 mm in the patient plane. The detected patient transformations at treatment deliveries can be used for a follow-up dose accumulation study. The purpose of the dose accumulation study is to allow the next treatment fraction to compensate for those regions that have been over-dosed and those regions that have been under-dosed.

3.3 Treatment Imaging Geometry

This section describes two IGRT imaging geometries investigated for this dissertation: Cone-Beam CT (3.3.1) and Nanotube Stationary Tomosynthesis (3.3.2).

3.3.1 Cone-beam CT (CBCT)

A CBCT is a rotational imaging system with a single radiation source and a planar detector, which may be mounted on a medical linear accelerator. This pair rotates by an angle of up to 2π during IGRT, taking projection images Ψ during traversal (Fig. 3.3.1(a)). A limited-angle rotation provides a shortened imaging time and lowered imaging dose. For example, a 5° rotation takes ~ 1 second. In my application, CBCT projections were acquired in a half-fan mode. Half-fan mode means that the imaging panel (40 cm width by 30 cm height, source-to-panel distance 150 cm) is laterally offset 16 cm to increase the CBCT reconstruction diameter to 46 cm. The method's linear operators are trained for projection angles over 360 degrees at 1 degree intervals beforehand at planning time. At treatment time my proposed methods will choose a learned regression that is closest to the current projection angle.

3.3.2 Nanotube Stationary Tomosynthesis (NST)

An NST is a stationary imaging system may be mounted on a medical linear accelerator so as to perform imaging without interfering with treatment delivery. As illustrated in Fig. 3.3.1(b), it consists of an arrangement of radiation sources arrayed around the treatment portal, together with a planar detector. The geometry thus is fixed and known beforehand. Firing the sources in sequence produces a sequence of projection images at different orientations. Each projection image requires ~ 200 ms.

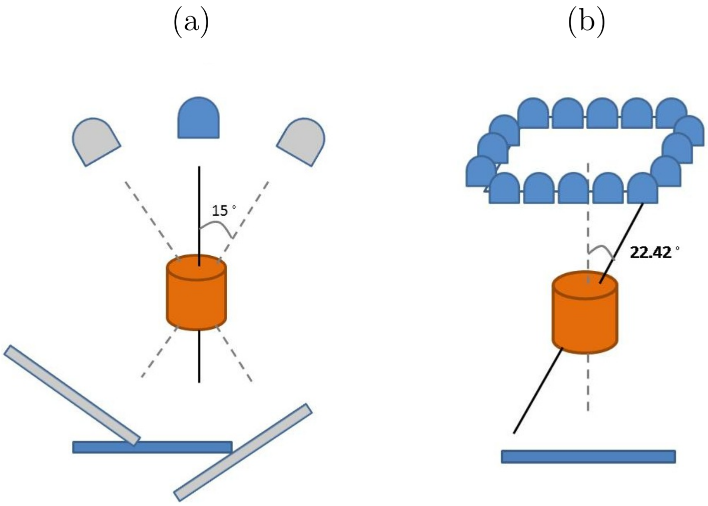

Figure 3.3.1: (a) Short arc CBCT geometry: rotational imaging system depicting a 30°arc. The image detector is laterally offset for half-fan acquisition. (b) The NST geometry: stationary sources array with angle $\theta = 22.42°$

3.4 Projection Intensity Pattern by a Local Normalization Scheme

X-ray scatter is a significant contributor to the cone-beam CT projections. However, the regression estimators of my proposed methods are not invariant to the projection intensity variations caused by x-ray scatter. Therefore, I implemented a normalization filter (3.4.1) and a subsequent histogram matching scheme (3.4.2) that when applied to both learning-time computed projections and registration-time target projections, generate commensurate intensities between these two images.

3.4.1 Local Gaussian Normalization

To account for variations caused by x-ray scatter, a 2D Gaussian-weighted normalization is performed on each pixel in the learning projections (Fig. 3.4.1(d)) and in the target projections (Fig. 3.4.1(b)). To calculate the normalized value $\Psi'(\mathbf{x};\theta)$ at pixel location $\mathbf{x} = (x_1, x_2)$ and projection angle θ, I subtract a Gaussian-weighted spatial mean $\mu'(x_1, x_2)$ from the raw pixel value $\Psi(x_1, x_2)$ and divide it by a Gaussian-weighted standard deviation $\sigma'(x_1, x_2)$.

$$\Psi'(x_1, x_2) = \frac{\Psi(x_1, x_2) - \mu'(x_1, x_2)}{\sigma'(x_1, x_2)} \tag{3.4.1}$$

$$\mu'(x_1, x_2) = \frac{\sum_{\xi=x_1-A}^{x_1+A} \sum_{\eta=x_2-B}^{x_2+B} [K(\xi, \eta; 0, w) \cdot \Psi(\xi, \eta)]}{\sum_{\xi=x_1-A}^{x_1+A} \sum_{\eta=x_2-B}^{x_2+B} K(\xi, \eta; 0, w)} \tag{3.4.2}$$

$$\sigma'(x_1, x_2) = \left(\frac{\sum_{\xi=x_1-A}^{x_1+A} \sum_{\eta=x_2-B}^{x_2+B} [K(\xi, \eta; 0, w) \cdot \Psi(\xi, \eta) - \mu'(x_1, x_2)]^2}{\sum_{\xi=x_1-A}^{x_1+A} \sum_{\eta=x_2-B}^{x_2+B} K(\xi, \eta; 0, w)} \right)^{\frac{1}{2}} \tag{3.4.3}$$

where $2A + 1$ and $2B + 1$, respectively, are the number of columns and rows in the averaging window centered at (x_1, x_2); the function K is an isotropic Gaussian with marginal standard deviation w. I choose A, B, and w to be appropriate values (discussed in Ch. 5.5.2) to perform a local Gaussian-weighted normalization for my target problem.

3.4.2 Histogram Matching

In order to correct the intensity spectrum differences between the normalized learning projection $\Psi'_{learning}$ and the normalized target projection Ψ'_{target}, a function F_ω of intensity to achieve non-linear cumulative histogram matching within a region of interest ω is applied. To avoid having background pixels in the histogram, the region ω is determined as that pixel set whose intensity values are larger than the mean value in the projection. That is, F_ω is defined by

$$F_\omega(H_f(\Psi'_{target})) \approx H_f(\Psi'_{learning}) \tag{3.4.4}$$

where H_f is the cumulative histogram profiling function. The histogram matched intensities $\Psi^{\star}_{target}$ (Fig. 3.4.1(c)) are calculated through the mapping:

$$\Psi^{\star}_{target} = \Psi'_{target} \circ F_{\omega} \tag{3.4.5}$$

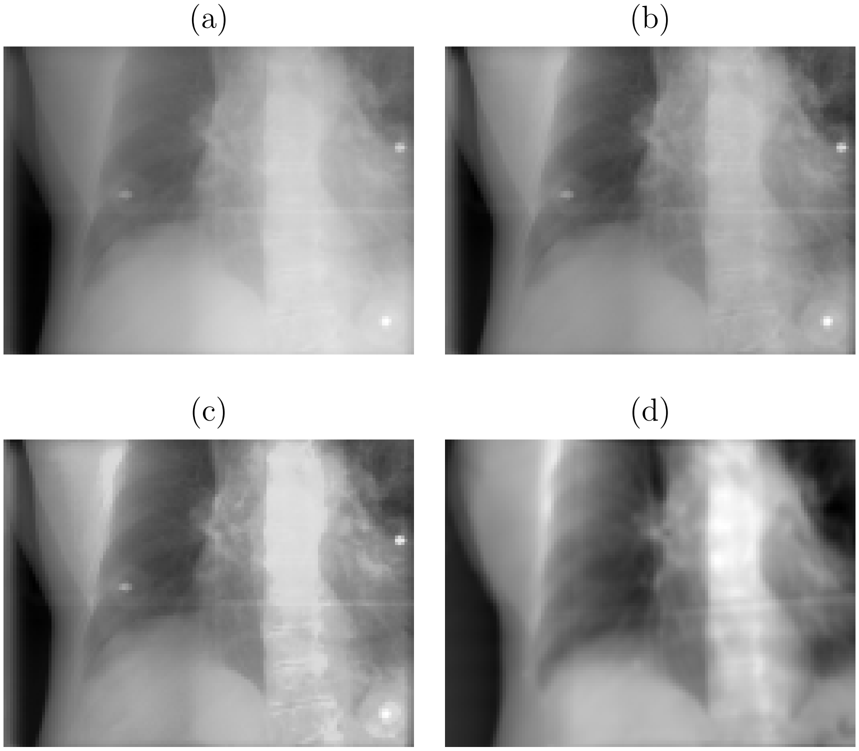

Figure 3.4.1: (a) A raw Cone-Beam CT (CBCT) projection (target projection), (b) a local Gaussian normalized CBCT projection (normalized target projection), (c) histogram matched CBCT projection (normalized and histogram matched target projection) and (d) a local Gaussian normalized DRR of a Fréchet mean CT (learning projection) from a lung dataset. As shown in the images, after normalization and histogram matching, the intensity contrast in the target projection becomes closer to that in the learning projection.

Chapter 4: Transformation Parameterization for IGRTs

To make the 2D/3D image registration robust, my methods limit the patient's transformation to a shape space. In order to have efficient registration, my methods also represent the patient's transformation with low-dimensional global parameters **C**. The following sections (4.1 and 4.2) detail the shape space formulation and parameterization for both rigid and non-rigid transformations.

4.1 Rigid Transformation

In IGRT sites where the patient's motion is mainly rigid (e.g., head and neck), the patient's motion can be modeled explicitly as the variation in the Euler's six dimensional rigid space:

$$\mathbf{C} = (t_x, t_y, t_z, r_x, r_y, r_z) \tag{4.1.1}$$

where t_x, t_y, t_z are the translation amounts in cm along the world's coordinate axes x, y, z, respectively; and r_x, r_y, r_z are the rotations in degrees about the image center, around the world coordinate axes x, y, and z, in succession.

4.2 Non-Rigid Transformation

In IGRT sites, where the patient's motion is mainly non-rigid (e.g., lung and abdominal IGRT), like others (Liu *et al.* (2010); Li *et al.* (2011a)) my methods model the non-rigid transformation (deformation) as a linear combination of a set of basis deformations calculated through principal component analysis (PCA). In lung and abdominal IGRT where the patient's deformation is dominated by respiration, a cyclically varying set of 3D images across respiration cycle $\{J_\tau\colon \Omega \subset \mathbb{R}^3 \text{ over time } \tau\}$ are available at treatment planning time. From these a mean image $\bar{J}$ and a set of deformations $\phi_\tau\colon \Omega \to \mathbb{R}^3$ between J_τ and $\bar{J}$ can be computed. The basis deformations are chosen to be the primary eigenmodes of the set of deformations $\{\phi_\tau\}$. The computed mean image $\bar{J}$ will be used as the reference image I throughout this dissertation. The following sections will detail the computation pipeline.

4.2.1 Deformation Shape Space and Mean Image Generation

In order to obtain a reference image that better represents the mean point in the patient's respiratory cycle, my methods compute a Fréchet mean image $\overline{J}$ that is an intrinsic mean image on the patient's respiratory manifold. The Fréchet mean image $\overline{J}$ can be computed by an *LDDMM* (Large Deformation Diffeomorphic Metric Mapping) framework (Beg *et al.* (2005)) from the cyclically varying set of 3D images $\{J_\tau$ over time $\tau\}$. The Fréchet mean, as well as the diffeomorphic deformations ϕ from the mean to each image J_τ, are computed using a fluid-flow distance metric d_{fluid} (Lorenzen *et al.* (2006)):

$$\overline{J} = \underset{J}{argmin} \sum_{\tau=1}^{N} d_{fluid}(J, J_\tau)^2 \tag{4.2.1}$$

$$= \underset{J}{argmin} \left(\sum_{\tau=1}^{N} \int_0^1 \int_\Omega ||v_{\tau,\gamma}(x)||^2 dx d\gamma + \frac{1}{\alpha^2} \int_\Omega ||J(\phi_\tau^{-1}(x)) - J_\tau(x)||^2 dx \right) \tag{4.2.2}$$

where $J_\tau(x)$ is the intensity of the pixel at position x in the image J_τ, $v_{\tau,\gamma}$ is the fluid-flow velocity field for the image J_τ in flow time γ , α is the weighting variable on the image dissimilarity, and $\phi_\tau(x)$ describes the deformation at the pixel location x:

$\phi_\tau(x) = x + \int_0^1 v_{\tau,\gamma}(x)d\gamma$.

The mean image $\overline{J}$ and the deformations ϕ_τ are calculated by gradient descent optimization. The set $\{\phi_\tau \text{ over } \tau\}$ can be used to generate the deformation shape space by the following statistical analysis.

4.2.2 Statistical Analysis

Starting with the diffeomorphic deformation set $\{\phi_\tau\}$, one can represent this diffeomorphic set by doing analysis on their initial momenta (Wang *et al.* (2007); Zhong and Qiu (2010); Niethammer *et al.* (2011)), e.g., on $v_{\tau,0}$ for registering J_τ to $\bar{J}$. However, the goal of the IGRT application is to do fast, and probably real-time, image registration. Although registration leveraging the analyzed space of the initial momenta will help to generate realistic diffeomorphic transformations constrained to the transformation space a priori, the registration evolution at treatment time is still time-consuming. Therefore, instead of analyzing the space of the initial momenta, my methods find a set of linear deformation basis functions ϕ_{pc}^i by doing PCA on the diffeomorphic set. The linear combination of the scores λ_τ^i (basis function weights) and the corresponding basis functions ϕ_{pc}^i yield a final transformation ϕ_τ in terms of these basis functions.

$$\phi_\tau = \overline{\phi} + \sum_{i=1}^{N} \lambda_\tau^i \cdot \phi_{pc}^i \tag{4.2.3}$$

A subset of n eigenmodes that capture 95% of the total variation are chosen, and they let the n basis function weights λ^i form the n-dimensional parameterization $\mathbf{C}$.

$$\begin{aligned} \mathbf{C} &= (c^1, c^2, \cdots, c^n) && (4.2.4) \\ &= (\lambda^1, \lambda^2, \cdots, \lambda^n) && (4.2.5) \end{aligned}$$

Chapter 5: Linear Regression Learning (CLARET)

This chapter begins with describing the general framework for the proposed 2D/3D registration methods (5.1). Section 5.2 details the proposed CLARET method that can do efficient registration using linear regression. The method's application for rigid registration involves a multi-scale learning scheme that is also detailed in Section 5.3. I describe the experimental setup in Section 5.4. In Section 5.5 I show CLARET's rigid (5.5.1) and non-rigid (5.5.2) 2D/3D registration results on synthetic and real test cases.

5.1 General 2D/3D Registration

The goal of the 2D/3D registration is to infer 3D transformations from 2D projections. I denote the projection intensity at pixel location $\mathbf{x} = (x_1, x_2)$ and projection angle θ as $\Psi(\mathbf{x}; \theta)$. $\Psi(\theta) \subset \mathbb{R}^{1\times\mathcal{P}}$ ($\mathcal{P}$ is the dimension of the 2D projection). The registration is formulated as an iterative process. Let I denote the 3D reference image and $I(t)$ denote the 3D image at iteration t. The estimated 3D image region's motion/deformation parameters $\hat{\mathbf{C}}(t)$: $\mathbb{R} \rightarrow \mathbb{R}^n$ (n is the number of parameters) define a geometric transformation $T(\hat{\mathbf{C}}(t))$: $\mathbb{R}^n \rightarrow \mathbb{R}^3$ in a space determined from one or more 3D images. $I(t)$ is $T(\hat{\mathbf{C}}(t))$ applied to $I(0)$: Eq. 5.1.1.

$$\begin{cases} I(t) &= I(0) \circ T(\hat{\mathbf{C}}(t)) \\ I(0) &= I \\ T(0) &= \mathbf{Id} \end{cases} \tag{5.1.1}$$

Id is the identity transformation.

The $\hat{\mathbf{C}}(t)$ are calculated by the estimated parameter updates $\Delta\hat{\mathbf{C}}(t)$: $\mathbb{R} \to \mathbb{R}^n$: Eq. 5.1.2.

$$\begin{cases} \hat{\mathbf{C}}(0) & = & \mathbf{0} \\ \hat{\mathbf{C}}(t) & = & \hat{\mathbf{C}}(t-1) + \Delta\hat{\mathbf{C}}(t) \end{cases} \tag{5.1.2}$$

The estimated parameter updates are obtained from the projection intensity residues $\mathbf{R} \subset \mathbb{R}^{1\times\mathcal{P}}$ ($\mathcal{P}$ is the dimension of the 2D projection) between the target 2D projections $\Psi(\mathbf{x};\theta)$ and the computed projections $\mathbf{P}(\mathbf{x}, I(t-1);\theta) \subset \mathbb{R}^{1\times\mathcal{P}}$ of the transformed 3D source image at iteration $t-1$: Eq. 5.1.3.

$$\mathbf{R}[\Psi(\mathbf{x};\theta), \mathbf{P}(\mathbf{x}, I(t-1);\theta)] = \Psi(\mathbf{x};\theta) - \mathbf{P}(\mathbf{x}, I(t-1);\theta) \tag{5.1.3}$$

After parameter estimation in each iteration, an image transformation (Eq. 5.1.1) is required in order to produce updated computed projections for the parameter estimation in the next iteration.

The projection operator simulates the imaging process. For example, according to the medical literature, to simulate a 3D image's x-ray projections from its 3D volume (DRRs) (Sherouse *et al.* (1990)), a ray-casting scheme can be used to compute the photon attenuation through a given imaging geometry (Fig. 5.1.1). Although Eq. 5.1.3 indicates a simple subtraction of the projection of the 3D image from the target projection, in actual clinical application one must apply additional processing to both the DRR and the real radiograph to remove the low varying additive scattering effect in the target projections.

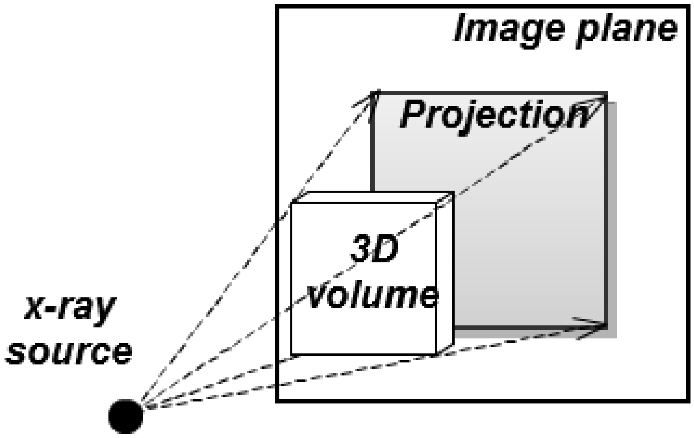

Figure 5.1.1: An x-ray projection is simulated by ray-casting on a 3D image volume. The dashed lines and arrows indicate the ray directions.

One way to obtain the estimated parameter updates $\Delta\hat{\mathbf{C}}(t)$ is by minimizing the sum of squared joint intensity residues $\mathbf{R}^{\dagger}$ at various angles: $\theta_1, \theta_2, ..., \theta_{\mathbf{A}}$ over the parameter updates $\Delta\mathbf{C}$.

$$\Delta\hat{\mathbf{C}}(t) = \underset{\Delta\mathbf{C}}{argmin} \left\| \mathbf{R}^{\dagger}[\Psi, \mathbf{P}(I(0) \circ T(\hat{\mathbf{C}}(t-1) + \Delta\mathbf{C}))] \right\|_{L_2}^{2} \qquad (5.1.4)$$

The joint intensity residues $\mathbf{R}^{\dagger} \subset \mathbb{R}^{1\times\mathcal{P}\cdot\mathbf{A}}$ are defined as a concatenation over the residues at available projection angles: $\mathbf{R}^{\dagger} = (\mathbf{R}_{\theta_1}, \mathbf{R}_{\theta_2}, \cdots, \mathbf{R}_{\theta_{\mathbf{A}}})$ with dimension$=1 \times \mathcal{P} \cdot \mathbf{A}$. Without parallelization, iterative computations to carry out this optimization are structurally slow as it involves the time-consuming evaluation of the functional Jacobian at each iteration. Moreover, the optimization may easily converge to a local minimum since the energy functional in Eq. 5.1.4 is not convex. See Section 5.5.2 for a detailed evaluation of the optimization-based approach.

5.2 Efficient Linear Approximation of ΔC

Instead of adapting the traditional approach (Markelj *et al.* (2012); Zollei *et al.* (2001); Li *et al.* (2010, 2011a); Russakoff *et al.* (2005, 2003); Weese *et al.* (1997); Fu and Kuduvalli (2008); Yao and Taylor (2003); Pickering *et al.* (2009); Sarrut and Clippe (2001); Rohlfing *et al.* (2002); Knaan and Joskowicz (2003)) as characterized in Eq. 5.1.4, I propose an alternative method (CLARET) to calculate the parameter updates $\Delta\mathbf{C} \subset \mathbb{R}^{1\times n}$ using a learned linear operator $\mathbf{W} \subset \mathbb{R}^{\mathcal{P}\cdot\mathbf{A}\times n}$ applied to projection intensities. At each iteration of the registration, the method estimates the motion/deformation parameter updates by applying a linear operator to the current joint intensity residue $\mathbf{R}^{\dagger} \subset \mathbb{R}^{1\times\mathcal{P}\cdot\mathbf{A}}$. That is,

$$\Delta\hat{\mathbf{C}}(t) = \mathbf{R}^{\dagger}[\Psi, \mathbf{P}(I(t-1))] \cdot \mathbf{W}, \tag{5.2.1}$$

where $t = 1, 2, \cdots, t_{max}$. Typically, $t_{max} \leq 10$ are satisfactory.

The computation in Eq. 5.2.1 only involves matrix multiplications by $\mathbf{W}$, computation of the projections $\mathbf{P}$, and subtractions (Eq. 5.1.3). Therefore, this registration can be computed efficiently. The calculation of the linear operator $\mathbf{W}$ involves a machine learning process described in detail in Section 5.3. Due to the machine learning process and the fast linear operation, the proposed CLARET method shows a more robust and faster registration than the optimization-based approach. See Section 5.5.2 for some comparisons.

5.3 Linear Regression Learning

From the motion/deformation shape space the CLARET method calculates the linear operator that correlates sampled shape parameters with the corresponding projection intensity residue vectors. I describe the regression learning that calculates the linear operators in Section 5.3.1 and an efficient sampling strategy in Section 5.3.2.

5.3.1 Learning Regressions from Residues to Shape Parameters

The regression learning is similar to that in the Active Appearance Model (AAM) (Cootes *et al.* (2001)). As detailed in Chapter 4, a collection of shape parameters $\{\mathbf{C}_\kappa \subset \mathbb{R}^{1\times n}$ over cases $\kappa\}$ are sampled for learning. Each case is formed by a selection of parameter settings. The training uses deviations from the reference image, such that $\Delta\mathbf{C} = \mathbf{C}_\kappa$. Linear regression is used to correlate the sampled parameters $\mathbf{C}_\kappa$ in the κ^{th} case with the co-varying projection intensity residue set $\{\mathbf{R}_{\kappa,\theta} \subset \mathbb{R}^{\mathbf{1}\times\mathcal{P}}$ over the projection angles $\theta\}$. $\mathbf{R}_{\kappa,\theta}(\mathbf{x})$ is computed as the intensity difference at pixel location $\mathbf{x} = (x_1, x_2)$ between the projection at angle θ of the mean image $I\colon \Omega \subset \mathbb{R}^3 \to \mathbb{R}$ (or an untransformed 3D image for the rigid case) and the projection of the image $I \circ T(\mathbf{C}_\kappa)$ transformed with the sampled model parameter $\mathbf{C}_\kappa$:

$$\mathbf{R}_{\kappa,\theta}(\mathbf{x}) = \mathbf{P}(\mathbf{x}, I \circ T(\mathbf{C}_\kappa); \theta) - \mathbf{P}(\mathbf{x}, I; \theta) \tag{5.3.1}$$

The method concatenates the residues at each projection angle to formulate a residue set in a vector $\mathbf{R}^\dagger_\kappa \subset \mathbb{R}^{\mathbf{1}\times\mathcal{P}\cdot\mathbf{A}} = (\mathbf{R}_{\kappa,\theta_\mathbf{1}}, \mathbf{R}_{\kappa,\theta_\mathbf{2}}, \cdots, \mathbf{R}_{\kappa,\theta_\mathbf{A}})$ and computes a linear regression for all cases $\kappa = 1, 2, \cdots, K$:

$$\begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_K \end{pmatrix} \approx \begin{pmatrix} \mathbf{R}^\dagger_1 \\ \mathbf{R}^\dagger_2 \\ \vdots \\ \mathbf{R}^\dagger_K \end{pmatrix} \cdot \mathbf{W} \tag{5.3.2}$$

The regression matrix $\mathbf{W} \subset \mathbb{R}^{\mathcal{P}\cdot\mathbf{A}\times n}$ that yield the best least square fitting to the training set can be computed via a pseudo-inverse (Peters and Wilkinson (1970)):

$$\mathbf{W} = (\mathbf{R}^{\dagger\intercal}\mathbf{R}^{\dagger})^{-1}\mathbf{R}^{\dagger\intercal}\mathbf{C} \tag{5.3.3}$$

where $\mathbf{R} \subset \mathbb{R}^{K\times\mathcal{P}\cdot\mathbf{A}}$ and $\mathbf{C} \subset \mathbb{R}^{K\times n}$.

5.3.2 Efficient Sampling

To provide adequate regression learning, $\mathbf{C}$ must be sufficiently sampled to capture all the shape variations. I have designed an efficient scheme to sample the shape parameters. Each shape parameter c^i is collected from the combinations of $\pm 3\sigma^i$ and 0 where σ^i is the standard deviation of the basis function weights λ^i observed at treatment planning time.

5.3.3 Linear Assumption for Iterative Estimation

The linear assumption between deformation parameters and intensity residues allows the method to estimate parameter differences not only from a fixed reference deformation but also from other deformations in the shape space. For example, assume there is a linear regression estimator $\mathbf{W}$ that has been trained to estimate deformation parameters $\mathbf{C} \subset \mathbb{R}^{K\times n}$ from a collection of K training deformation parameter values. Now consider two training parameter values $\mathbf{C}_p \subset \mathbb{R}^{1\times n}$ and $\mathbf{C}_q \subset \mathbb{R}^{1\times n}$. Recall that the linear regression estimator $\mathbf{W} \subset \mathbb{R}^{\mathcal{P}\cdot\mathbf{A}\times n}$ is designed to estimate their values from their intensity residues, $\mathbf{R}_p^{\dagger} \subset \mathbb{R}^{1\times\mathcal{P}\cdot\mathbf{A}}$ and $\mathbf{R}_q^{\dagger} \subset \mathbb{R}^{1\times\mathcal{P}\cdot\mathbf{A}}$ respectively, from the reference DRR (Eq. 5.3.4 and Eq. 5.3.5). $\mathbf{R}_p^{\dagger} = \mathbf{P}(I \circ T(\mathbf{C}_p)) - \mathbf{P}(I)$, and $\mathbf{R}_q^{\dagger} = \mathbf{P}(I \circ T(\mathbf{C}_q)) - \mathbf{P}(I)$.

$$\mathbf{C}_p = \mathbf{R}_p^{\dagger} \cdot \mathbf{W} \tag{5.3.4}$$

$$\mathbf{C}_q = \mathbf{R}_q^\dagger \cdot \mathbf{W} \tag{5.3.5}$$

Notably, the regression result that estimates deformations from the mean to the training images can estimate the update in $\mathbf{C}_q$ to produce $\mathbf{C}_p$ from their intensity residue difference: $\mathbf{R}_p^\dagger - \mathbf{R}_q^\dagger$ as well. See Eq. 5.3.6.

$$\mathbf{C}_p - \mathbf{C}_q = \{\mathbf{R}_p^\dagger - \mathbf{R}_q^\dagger\} \cdot \mathbf{W} \tag{5.3.6}$$

With this property the method can perform efficient regression learning. It only needs to train deformation differences from a single reference deformation ($\mathbf{C} = 0$), but the learned regression can estimate deformation update from any deformations as well.

If the underlying deformation-to-intensity-residues relationship is not nearly linear, the error will propagate. Other proposed methods described in Ch. 6 and Ch. 7 are trying to relax this strong assumption of linearity.

5.4 Experimental Setup

The experimental setups for evaluating the method are described in this section. In IGRT the 3D image I is the planning CT, and the target projection images Ψ are treatment-time imaging kV projections. In particular, the kV projections are produced by (1) a rotational *CBCT* (Cone-beam CT) imager or (2) the stationary *NST* (Nanotube Stationary Tomosynthesis) imager specified in Maltz *et al.* (2009). See Sections 3.3.1 and 3.3.2 for the geometries. The experimental setups for head-and-neck and lung IGRTs are detailed in Sections 5.4.1 and 5.4.2, respectively.

5.4.1 Head-and-neck IGRT

In head-and-neck IGRT, the geometric differences of the skull between planning time and treatment time can be represented by a rigid transformation. Therefore, at treatment planning time, CLARET samples clinically feasible variations (±2 cm in translations, ±5° in rotations) in the Euler 6-space $\mathbf{C}$ to capture the treatment-time patient's motions. With a single planning CT I of the patient, the computed learning projections $\mathbf{P}(I \circ T(\mathbf{C}); \theta)$ are generated by transformation of the feasible variations $T(\mathbf{C})$ and projection from a given angle θ of the transformed 3D volume $I \circ T(\mathbf{C})$.

In the registration, CLARET iteratively applies the linear operators to estimate the rigid transformation from the 2D intensity residues formed by the difference between the normalized target projections $\Psi^{\star}$ and the normalized projections computed from the currently estimated rigid transformation applied to the planning-time 3D image.

5.4.2 Lung IGRT

A consideration in lung IGRT is that respiratory motion introduces non-rigid transformations. At treatment planning time, 10-phase RCCT set (Respiratory-Correlated CTs) collected at planning time serve as the cyclically varying 3D images $\{J_\tau$ over the phase $\tau\}$. This image set is used to generate the deformation shape space $\mathbf{C}$. From these RCCTs, a Fréchet mean image $\overline{J}$ and its deformations ϕ_τ to the corresponding images J_τ are calculated via an LDDMM framework 4.2.2. Fig. 5.4.1(c) shows an example respiratory Fréchet mean image.

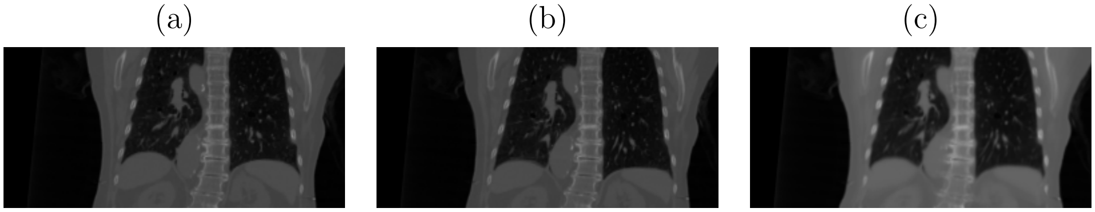

Figure 5.4.1: Coronal slices of (a) Respiratory-Correlated CT (RCCT) at the End-Expiration (EE) phase (b) RCCT at the End-Inspiration (EI) phase and (c) Fréchet mean CT generated via the Large Deformation Diffeomorphic Metric Mapping framework from the RCCT dataset.

The deformation basis functions ϕ_{pc} are then generated by PCA on the deformation set $\{\phi_\tau \text{ over phase } \tau\}$. Liu *et al.* (2010); Chou *et al.* (2013) has shown that a shape space with three eigenmodes adequately captures 95% of the total variance of respiratory variations experienced at treatment time. Fig. 5.4.2 shows the first two principal deformation basis functions.

To generate feasible variations in the deformation space $\mathbf{C}$ for learning the linear operator $\mathbf{W}$, CLARET samples parameters within three standard deviations of the means of the basis function weights derived from the RCCT image set. From the Fréchet mean image the computed projections $\mathbf{P}(\mathbf{x}, I \circ T(\mathbf{C}); \theta)$ are generated by

(1) transformation based on the feasible variations $T(\mathbf{C})$ and (2) projection from a given angle θ to the transformed 3D volume $I \circ T(\mathbf{C})$.

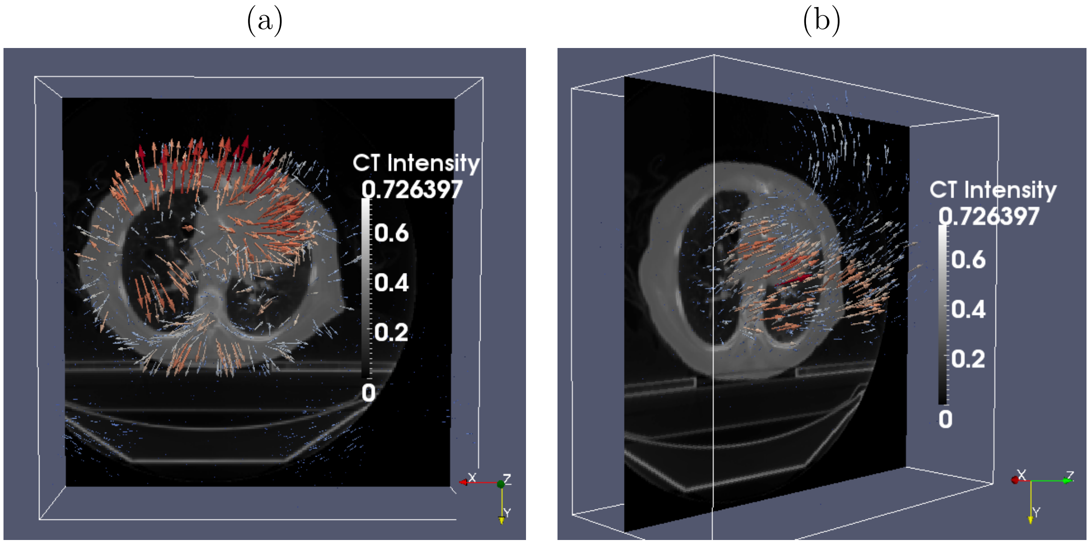

Figure 5.4.2: The (a) first and (b) second principal deformation basis functions analyzed from a lung RCCT dataset. Colored lines indicate heated body spectrum presentations of the deformation magnitudes. As shown in the images, the first principal motion consists of anterior-posterior expansion and contraction of the lung, and the second principal motion is along the superior-inferior direction. Compass in the figure: $\overrightarrow{X}$: Left to Right (LR); $\overrightarrow{Y}$: Anterior to Posterior (AP); $\overrightarrow{Z}$: Superior to Inferior (SI).

Just prior to treatment, the Fréchet mean image obtained at planning time is rigidly registered to the CBCT for correcting patient position. During treatment with planar imaging, CLARET iteratively applies the linear operators $\mathbf{W}$ to estimate the weights $\mathbf{C}$ on the basis functions ϕ_{pc} from current 2D intensity residues. The residues are formed by the difference between the normalized and histogram matched target projections $\Psi^{\star}$ (Fig. 3.4.1(c)) and the normalized projections (Fig. 3.4.1(d)) computed from the presently estimated deformation applied to the Fréchet mean image.

5.5 Results

Sections 5.5.1 shows rigid registration using the NST imaging system for the head-and-neck IGRT. Section 5.5.2 shows non-rigid registration using projection images from CBCT scans acquired with the rotational imaging system for lung intra-treatment IGRT. Section 5.5.2 compares registration accuracy and efficiency of CLARET and an optimization-based approach.

5.5.1 Rigid Registration Results

CLARET's rigid registration is tested by synthetic treatment-time projections and by real phantom projections, as described in Sections 5.5.2 and 5.5.2, respectively. The registration quality was measured by the mean absolute error (MAE) and mean target registration error ($mTRE$). The MAE in any of the parameters of $\mathbf{C}$ is the mean, over the test cases, of the absolute error in that parameter. The mTRE for a test case is the mean displacement error, over all voxels in a $16 \times 16 \times 16$ cm^3 bounding box (the probable tumor region) centered on the pharynx in the planning CT I.

$$mTRE := \left(\frac{\sum_{i=1}^{\chi} (\mathbf{y}_i \circ T(\mathbf{C}_{true}) - \mathbf{y}_i \circ T(\mathbf{C}_{est}))^2}{\chi} \right)^{\frac{1}{2}} \tag{5.5.1}$$

where χ is the number of pixels in the probable tumor region, $\mathbf{y}_i = (y_1, y_2, y_3)$ is the tuple of the i^{th} voxel position, and $\mathbf{C}_{true}$, $\mathbf{C}_{est}$ are the true and the estimated transformation parameters, respectively.

Synthetic Treatment Projections

I used noise-added DRRs (digitally reconstructed radiographs) of target CTs as the synthetic treatment-time projections. The DRRs (Fig. 3.4.1(a)) were generated to simulate the NST projections with dimension 128×128 and pixel spacing 3.2 mm (Fig.

5.1.1). The target CTs were transformed from the patient's planning CT by taking normally distributed random samples of the translation and rotation parameters within the clinical extent: ±2 cm and ±5°, respectively. The planning CTs have a voxel size of 1.2 mm lateral, 1.2 mm anterior-posterior, and 3.0 mm superior-inferior.

Zero mean, constant standard deviation Gaussian noise was added to the DRRs to generate the synthetic projections. The standard deviation of the noise was chosen to be 0.2 × (mean bony intensity - mean soft tissue intensity). This noise level is far higher than that produced in the NST system. An example synthetic projection is shown in Fig. 3.4.1(b).

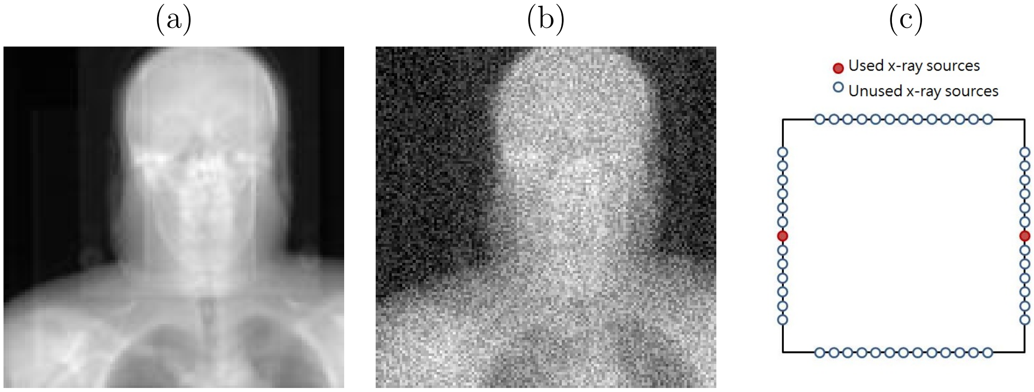

Figure 5.5.1: (a) A raw DRR from a x-ray source in the NST (b) DRR with Gaussian noise added (c) the NST geometry of two opposing x-ray sources

I first studied how many projection images are needed for CLARET's learning to obtain sub-voxel accuracy. The results on 30 synthetic test cases of a head-and-neck dataset, displayed in Fig. 5.5.2, show that two projection images are sufficient to achieve sub-CT-voxel accuracy. Fig. 5.5.2 also shows the method's accuracy improves with more projections. However, the redundant projections may contribute error in the parameter estimation. Therefore, the 4-projection geometry (Fig. 5.5.2), which used the middle x-ray source on each imaging bank, produced the sufficient and necessary number of projections to capture the simulated rigid motions. Fig. 3.4.1(c) shows the geometry of the two opposing x-ray sources that generated the two projection images in the study. The choice of opposing sources is chosen such that the maximum angle between images (22.5°) is formed with the NST system.

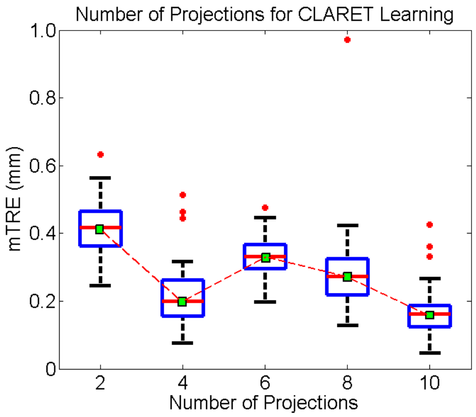

Figure 5.5.2: Boxplot results of errors in varying the number of projections used. Red dots are the outliers. Projections of equally spaced sources were used.

Table 7.1 shows the statistics of the errors in each rigid parameter from 90 synthetic test cases generated from three patients' planning CTs (30 cases for each CT). The CLARET registration used only the two opposing NST projection images (Fig. 3.4.1(c)).

(mm; °)	Tx	Ty	Tz	Rx	Ry	Rz	mTRE
MAE	0.094	0.302	0.262	0.1489	0.0248	0.1540	0.524
SE	0.008	0.022	0.075	0.011	0.001	0.030	0.076

Table 5.1: Mean absolute errors (MAE) and standard error (SE) of the absolute errors of the six rigid parameters obtained from 90 synthetic test cases in which CLARET registration used two synthetic NST projection images.

Real Treatment Projections

CLARET's rigid registration was also tested on a head phantom dataset. NST projections (dimension: 1024×1024; pixel spacing: 0.4 mm) of the head phantom were downsampled to dimension 128×128 with a pixel spacing of 3.2 mm (Fig. 5.5.3(a)). The dimension of the planning CT is $512 \times 512 \times 96$ with a voxel size of 3.43 mm^3. The comparison standard was obtained by rigidly registering the combined set of 52 NST projections to the planning CT by the l-BFGS optimization (Nocedal (1980)) of the similarity metric in projection space. This is not a good validation, but with no ground truths available optimization using all the projection images is our best try. Also, results in Frederick *et al.* (2010) suggest that 2D/3D registration accuracy is higher than limted-angle-reconstructed-3D/3D registration accuracy for the NST geometry. The initial mTRE over the head region was 51.8 mm. To deal with this exceptionally large initial deviation[1], CLARET trained regressions at 4 scales: $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3,$ and $\mathbf{W}_4$. At the s^{th} scale of training ($s = 1, 2, 3,$ and 4), each rigid parameter is collected from the combinations of $\mathcal{E} \cdot (5 - s)/4$ and 0 where $\mathcal{E} \subset \mathbb{R}$ is an extreme value selected for capturing this large setup deviation: $t_x, t_y, t_z = 80$ mm; $t_x, t_y, t_z = 20$ degrees. In the registration stage the calculated multiscale linear operators are applied sequentially, from $\mathbf{W}_1$ to $\mathbf{W}_4$, to give new estimations of the rigid parameters from large to small scale. The rationale behind this is that CLARET's regression accuracy is tied to the locality of the transformation: one should learn different regressions for different transformation local neighborhoods to have more accurate estimation. For rigid registration that has to deal with a high range of patient movements, e.g., from 5 cm to 1 mm, this multi-scale learning is used to estimate parameters in a coarse-to-fine fashion. For non-rigid registration, the iterative version of CLARET did not apply this

---

[1] Clinical initial setup deviations for head-and-neck IGRTs are 8.88 ± 2.53 mm. (Zeidan *et al.* (2007))

multi-scale learning. However, later in Ch. 7 I propose a noniterative and locally-linear regression learning framework that learns more accurate regression estimators for each deformation neighborhood.

With the 4-scale training ($S = 4$), CLARET obtained a sub-CT-voxel mTRE of 3.32 mm using only two projections in 5.81 seconds. It was computed on a 16-core laptop GPU (NVIDIA GeForce 9400m) where the parallelization is limited. A factor of 32 speed-up (˜0.18 seconds per registration) can be expected when using a 512-core GPU. As shown in Fig. 5.5.3(b) and 5.5.3(c), CLARET accuracy improves with increased number of projections and scales in the multi-scale learning process. The registration time is approximately linear with the number of projections used.

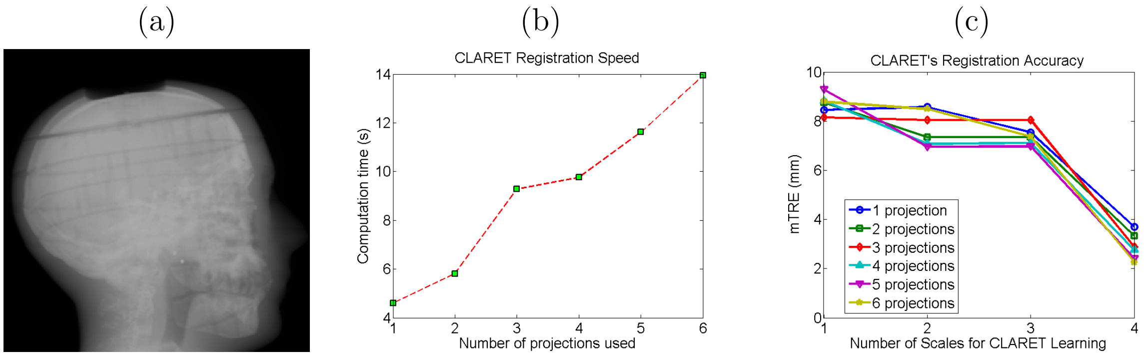

Figure 5.5.3: (a) One of the test NST projections of a head phantom. (b) Time plots and (c) error plots of CLARET's registrations on a real head-and-neck phantom dataset. Registrations were accelerated on a 16-core laptop GPU (NVIDIA GeForce 9400m).

5.5.2 Non-rigid Registration Results

CLARET's non-rigid registration is tested with synthetic and real patient cone-beam projections, as described in Sections 5.5.2 and 5.5.2, respectively. In Section 5.5.2, CLARET's registration accuracy and efficiency are also compared to the state-of-the-art method (Li *et al.* (2011a)) that uses the optimization approach. RCCT datasets (CT dimension $512 \times 512 \times 120$; voxel size 1 mm lateral × 1 mm anterior-posterior × 2.5

mm superior-inferior) were generated with an 8-slice scanner (LightSpeed, GE Medical Systems) by acquiring multiple CT images for a complete respiratory cycle at each couch position while recording patient respiration (Real-time Position Management System, Varian Medical Systems). The CT projections were retrospectively sorted (GE Advantage 4D) to produce 3D images at 10 different respiratory phases.

Synthetic Treatment Projections

DRRs of the target CTs were used as the synthetic treatment-time projections. The DRRs were generated to simulate projections from a rotating kV imaging system (Section 3.3.1) mounted on the gantry of the medical accelerator (TrueBeam, Varian Medical Systems). The target CTs were deformed from the patient's Fréchet mean CT by taking normally distributed random samples of the coefficients of the first three PCA-derived deformation eigenmodes of the patient's RCCT dataset (Section 4.2).

For each of the 10 CLARET registrations, I used a *single* simulated coronal projection (dimension 128×96; pixel spacing 3.10 mm) at angle 14.18° (Fig. 3.4.1(d)) as input. The registration quality was then evaluated by measuring the 3D tumor centroid difference between the CLARET-estimated CT and the target CT. 3D tumor centroids were calculated from active contour (geodesic snake) segmentations (Yushkevich *et al.* (2006)). As shown in Table 5.2, after registration CLARET reduces the centroid error more than 85%.

Case #	1	2	3	4	5	6	7	8	9	10
Before	8.2	21.3	21.8	20.1	9.9	10.2	10.9	15.7	14.9	19.9
After	1.3	0.8	1.5	3.3	0.8	1.3	0.5	1.6	2.1	2.7
% reductions	85	96	93	84	92	87	95	90	86	86

Table 5.2: 3D tumor centroid error (mm) before and after CLARET's registration for the 10 randomly generated test cases.

CLARET's registration quality was also studied in terms of average *DVF* (Displacement Vector Field) error over all cases and all CT voxels versus different angular spacings used in learning. Registrations using two projections with four different angle separations were tested by 30 randomly generated test cases. Fig. 5.5.4(a) shows that the average DVF error is small with appropriately large angular separations. However, tumor motion or respiratory motion may not be visible or inferable in projections from certain angles. For example, the tumor may be obscured by denser organs (e.g., mediastinum). In Fig. 5.5.4(a) the respiration motion may not be inferable from the projection at 9.68° resulting in a larger error in the parameter estimation.

I also studied CLARET's registration quality by measuring the average DVF error versus the number of projections used for learning. For each number of projections, I generated 30 random test cases. Fig. 5.5.4(b) shows no particular trend. As a result, I used a single projection to test CLARET's non-rigid registration for the real patient data in the next section. Also, giving the comparable number of outliers, the comparisons are valid for both plots.

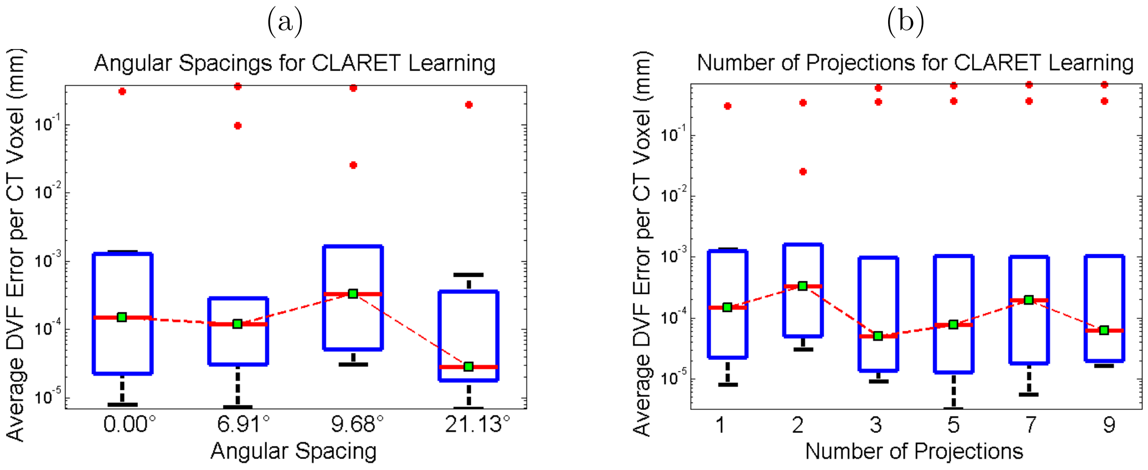

Figure 5.5.4: Boxplots of average displacement vector field errors when varying (a) the angular spacing and (b) the number of projections used for CLARET's non-rigid registration. Red dots are the outliers. In (a), two projections for each test were used. For the zero-degree test case, only one projection was used. In (b), DRRs spanning 9.68° about 14.18° were used in each test. The single projection was tested at 14.18° (see Fig. 3.4.1(d)).

Real Treatment Projections

CLARET was also tested using 5 lung patient datasets consisting of projections from patient CBCT scans acquired with the rotational imaging system (Section 3.3.1). CLARET 2D/3D registration used a single coronal CBCT projection at angle 14.18° and downsampled (original dimension 1024 × 768 and pixel spacing 0.388 mm) to 128 × 96 with 3.10 mm pixel spacing (Fig. 3.4.1(a)). Separate registrations were done using projections at the *EE* (End-Expiration) and *EI* (End-Inspiration) phases. Prior to 2D/3D registration, the Fréchet mean image was rigidly registered to the patient's CBCT image so as to align the vertebral anatomy in both images. I measured the difference in 3D tumor centroid position (Fig. 5.5.6(a)) between the CLARET-estimated CT and the reconstructed CBCT at the same respiratory phase as the projection used in the 2D/3D registration. The RMS window width was set to 32.0 mm for the Gaussian normalization for this imaging geometry, which was predetermined to yield the smallest 3D centroid error in one lung dataset (Fig. 5.5.5). Future studies should check whether this window size is also best for other datasets.

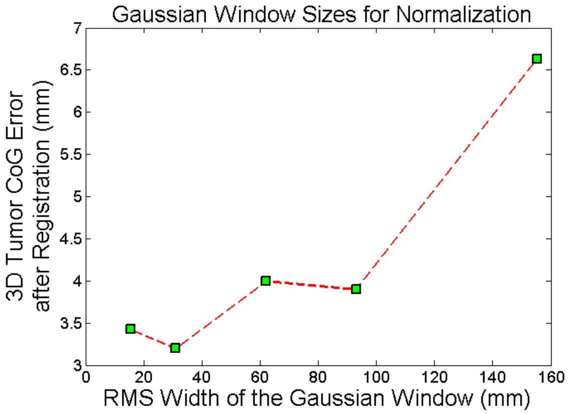

Figure 5.5.5: 3D tumor centroid error plots on a lung dataset for varying width of the Gaussian window used for CLARET's local Gaussian normalization.

The results shown in Table 5.3 suggest a consistency in registration quality between the synthetic image tests and real projection image tests. The mean and standard deviation of 3D tumor centroid errors following 2D/3D registration are 2.66 mm and 1.04 mm, respectively.[2] The average computation time is 2.61 seconds on a 128-core GPU, NVIDIA GeForce 9800 GTX. A factor of four speed-up (to 0.65 seconds) can be expected when using a 512-core GPU for acceleration.

The clinical goal is to improve tumor localization during treatment using CLARET. Assuming a mean lung tumor motion extent of about 10 mm, the standard deviation is about one-third of the motion extent, or 3 mm. In order to improve on current clinical practice (i.e., no image guidance during treatment) an mTRE of 2 mm or less is

---

[2] The errors include an uncertainty in tumor position in the CBCT projections, owing to variability in the manual segmentations in the CBCT reconstructions, and residual tumor motion within the EE and EI phase intervals. Based on repeatability measurements of the manual segmentations, the standard deviation in manually determined tumor 3D position is 1 mm.

desirable. Furthermore, since most of the motion is in the superior-inferior direction, it is desirable to achieve an mTRE of 2 mm or less in that direction. Our results show that CLARET achieves the clinically desired accuracy: the mean and standard deviation of the 2D tumor centroid error on the patient plane (left-right and superior-inferior directions) after registration is 1.96 mm and 1.04 mm, respectively. CLARET reduces positional errors in directions along the plane of the projection more than in the out-of-plane direction. As shown in Table 5.3, except in cases from patient #1 most of the percent 2D error reductions in the imaging plane (which was coronal) are larger than the percent error reductions in the out-of-plane direction. This is expected because 2D/3D registration with a single projection is more sensitive to tumor displacements in the image plane but less sensitive to scale changes due to out-of-plane displacements.

Fig. 5.5.6(b) shows the 3D meshes of the tumors in the Fréchet mean CT, the CBCT at EE, and the estimated CT of a lung dataset for visual validation. As shown in the figure, the tumor position in the CLARET-estimated CT is superior to that in the mean image, as expected physiologically for the EE phase. Fig. 5.5.7 shows the same 3-space lines in the mean CT, the reconstructed CBCT at the EE phase and the CLARET-estimated CT of a lung dataset. The intersections of the lines with the tumor centroid in the CBCT are in better agreement with the CLARET-estimated CT than with the mean CT, indicating that CLARET can accurately locate the tumor in the plane of the projection (coronal plane) and corroborating the results of Table 5.3.

Patient #	e^{3D}_{EE} (mm)	e^{2D}_{EE} (mm)	$e^{\perp}_{EE}$ (mm)	e^{3D}_{EI} (mm)	e^{2D}_{EI} (mm)	$e^{\perp}_{EI}$ (mm)	Time (s)
1	2.27 (7.96,72%)	2.07 (4.16,50%)	0.93 (6.78,86%)	5.26 (8.03,34%)	4.71 (6.80,31%)	2.34 (4.27,45%)	1.94±0.74
2	3.20 (9.70,67%)	2.23 (9.18,76%)	2.29 (3.13,26%)	2.85 (7.45,62%)	1.64 (6.77,75%)	2.33 (3.10,24%)	3.99±1.99
3	1.32 (1.47,10%)	1.32 (1.47,10%)	0.00 (0.00,0%)	2.03 (3.63,44%)	1.86 (3.49,47%)	0.81 (0.99,18%)	2.45±0.15
4	2.77 (10.17,73%)	1.51 (9.67,84%)	2.32 (3.14,26%)	2.31 (5.53,58%)	1.52 (5.17,71%)	1.73 (1.96,11%)	1.96±0.02
5	2.24 (3.52,36%)	1.91 (3.46,45%)	1.17(0.64,-82%)	2.40 (3.89,38%)	0.83 (3.16,74%)	2.25 (2.26,0%)	2.76±1.26

Table 5.3: Computation time and tumor centroid errors (mm) for 5 patient datasets. e^{3D}: 3D errors; e^{2D}: errors on the patient plane (coronal plane); $e^{\perp}$: errors orthogonal to the patient plane (out-of-plane direction). Subscripts of e indicate the respiration phases: EE or EI. Numbers inside the parentheses are, respectively, the tumor centroid error (mm) before CLARET registration (using the Fréchet mean image), and percentage error reduction after registration.

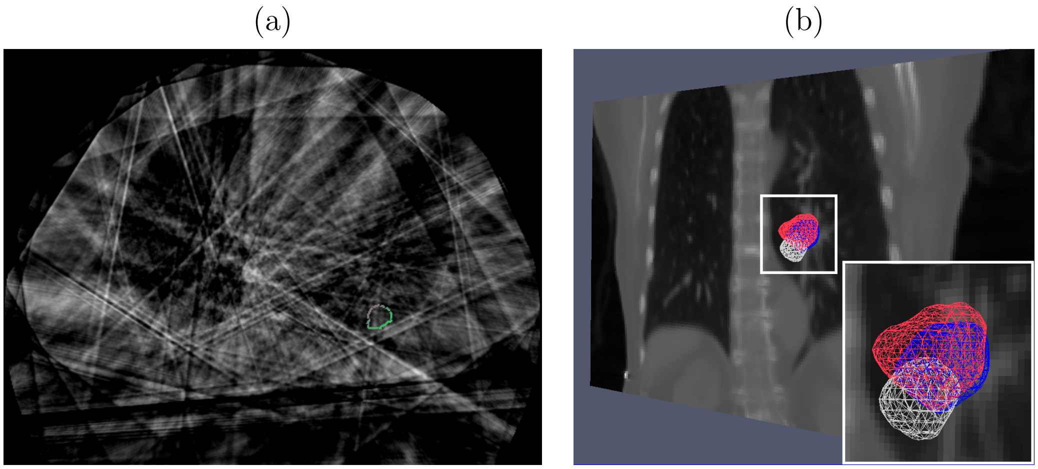

Figure 5.5.6: (a) Transaxial image with manual segmented tumor contours in the reconstructed CBCT at one respiratory phase of a lung dataset (patient 3). The contours were used for 3D centroid calculation. (b) Tumor meshes in the Fréchet mean CT (white), in the target CBCT at the EE respiratory phase (blue) and in the CLARET-estimated CT (red) of a lung dataset (patient 2). The background is a coronal slice of the mean CT for illustration. The overlap between the estimated and the target tumor meshes indicates a good registration.

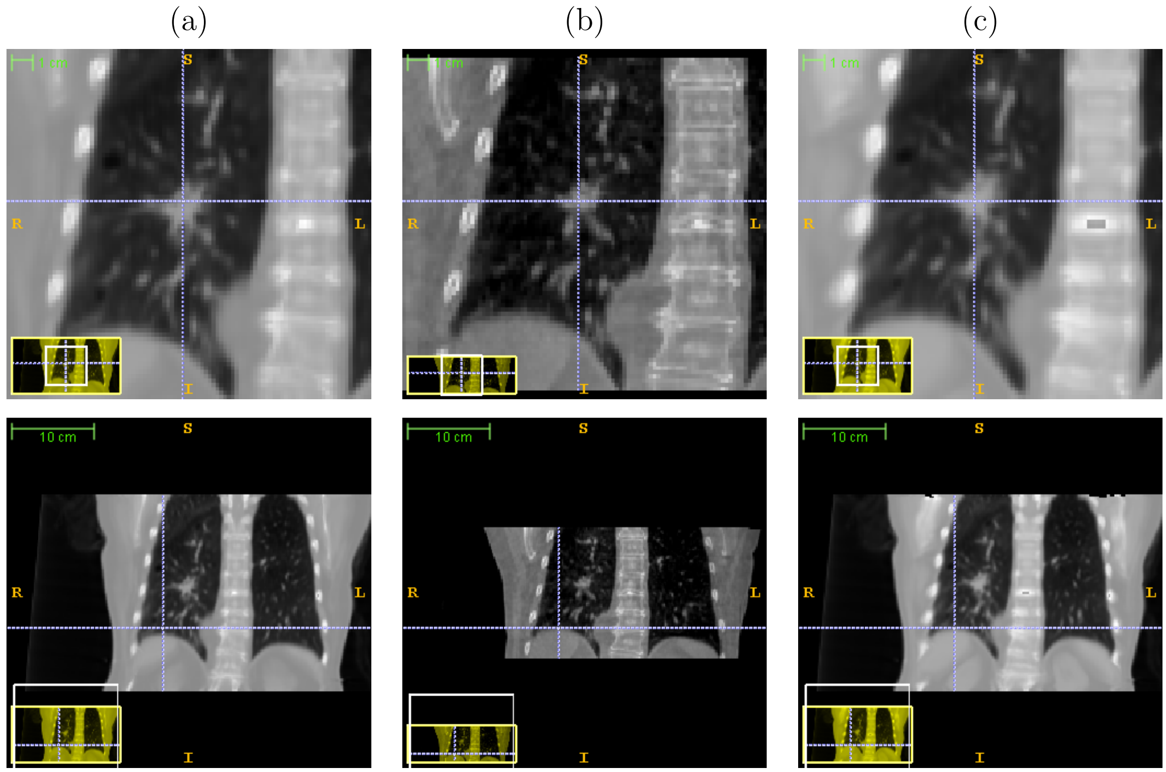

Figure 5.5.7: The same 3-space lines in (a) the mean CT, (b) the reconstructed CBCT at the EE phase and (c) the estimated CT of the same lung dataset used in figure 5.5.6(b). Upper row: lines indicate the tumor centroid in the CBCT at the EE phase; lower row: lines indicate the diaphragm contour in the CBCT at the EE phase.

Comparison to an Optimization-based Registration Method

CLARET's registration accuracy and efficiency have been compared to an optimization-based method similar to that in the state-of-the-art method (Li *et al.* (2011a)). The optimization-based method I implemented optimizes Eq. 5.1.4 (with $\rho = 2$) using the l-BFGS quasi-Newton algorithm (Nocedal (1980)). To make fair comparisons, I used the same deformation shape space, the same initializations, the same GPU acceleration for the projection operator $\mathbf{P}$, and the same testing datasets.[3]

For the comparisons, 30 randomly sampled synthetic deformations were used as the test cases for each of the five lung patients. The deformations were sampled randomly within ±3 standard deviations of deformations observed in the patient's RCCT. For each test case, a single coronal CBCT projection (dimension: 1024×768 downsampled to dimension: 128×96) was simulated from the deformed Fréchet mean CT as the target projection. Both methods were initialized with the realistic Fréchet mean image with no deformation: $\hat{\mathbf{C}}(0) = \mathbf{0}$ in Eq. 5.1.2.

For CLARET, each training deformation parameter c^i $(i = 1, 2, 3)$ was collected from the combinations of $\pm 3\sigma^i$, $\pm 1.5\sigma^i$, and 0 where σ^i is the standard deviation of the i^{th} eigenmode weights observed in the patient's RCCT. Therefore, for sampling on three eigenmodes, a total of 125 training deformations are sampled for each patient.

Registration accuracy was measured by the average registration error distance over the lung region. As Fig. 5.5.8 shows, CLARET yielded more accurate results than the l-BFGS optimization-based registration in almost every test case in all five patients. Table 5.4 showed statistical comparisons of the registration accuracy. The maximum error produced by CLARET among the $30 \times 5 = 150$ test cases is only 0.08 mm where the maximum error produced by l-BFGS is 13.15 mm, which is 164 times higher than CLARET. The smaller median error and error standard deviation also showed that CLARET is more accurate and more robust than the l-BFGS optimization-based approach.[4]

In term of registration speed, Fig. 5.5.9 showed that CLARET is faster than l-BFGS in every test case and has relatively small variation in speed. The statistical results

---

[3] Despite this my implementation is not completely the same as the method in Li *et al.* (2011a), but I had no access to the implementation in Li *et al.* (2011a).

[4] Note that the l-BFGS optimization-based approach also yields accurate registration with a mean mTRE of 0.54 mm.

shown in Table 5.5 indicate that the longest registration time produced by CLARET is still shorter than the shortest time produced by l-BFGS.

As our results show, in my implementations CLARET is more robust, accurate, and faster than the l-BFGS optimization.

mTRE (mm)	min.	max.	median	mean	std
CLARET	$1.1e^{-5}$	0.08	$2.3e^{-4}$	$1.5e^{-3}$	$7.4e^{-3}$
l-BFGS	$2.0e^{-4}$	13.15	$8.8e^{-3}$	0.54	2.01

Table 5.4: Registration accuracy (mTRE) statistics on the five patient data: CLARET vs. the l-BFGS optimization. std=standard deviation

time (s)	min.	max.	median	mean	std
CLARET	0.94	5.15	1.73	1.95	0.74
l-BFGS	5.29	78.73	19.30	23.76	14.41

Table 5.5: Registration time statistics on the five patient data: CLARET vs. the l-BFGS optimization. std=standard deviation

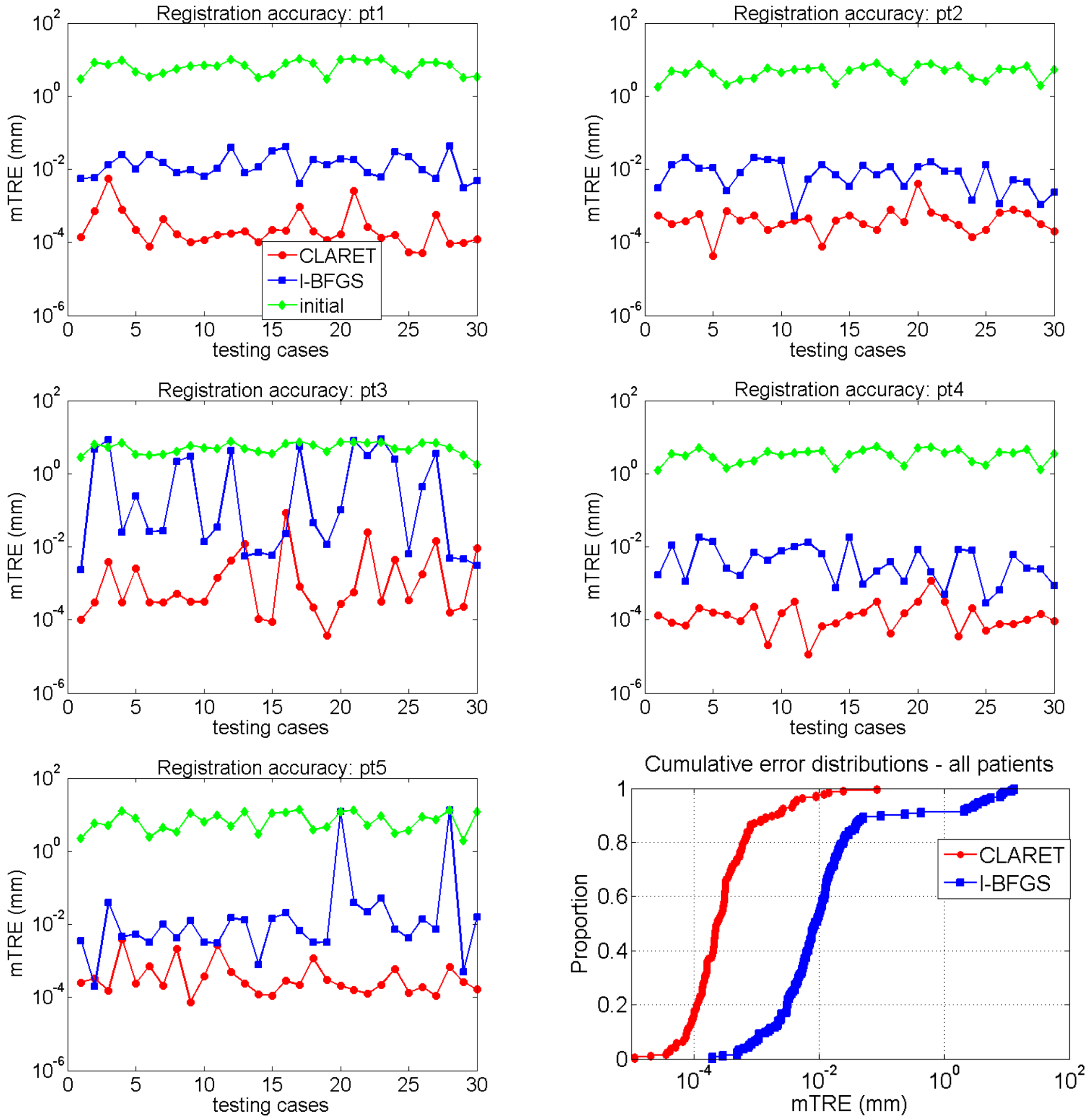

Figure 5.5.8: Mean target registration error (mTRE) on test data generated from five patients (pt1-pt5): CLARET vs. the l-BFGS optimization vs. the initial error before registration.

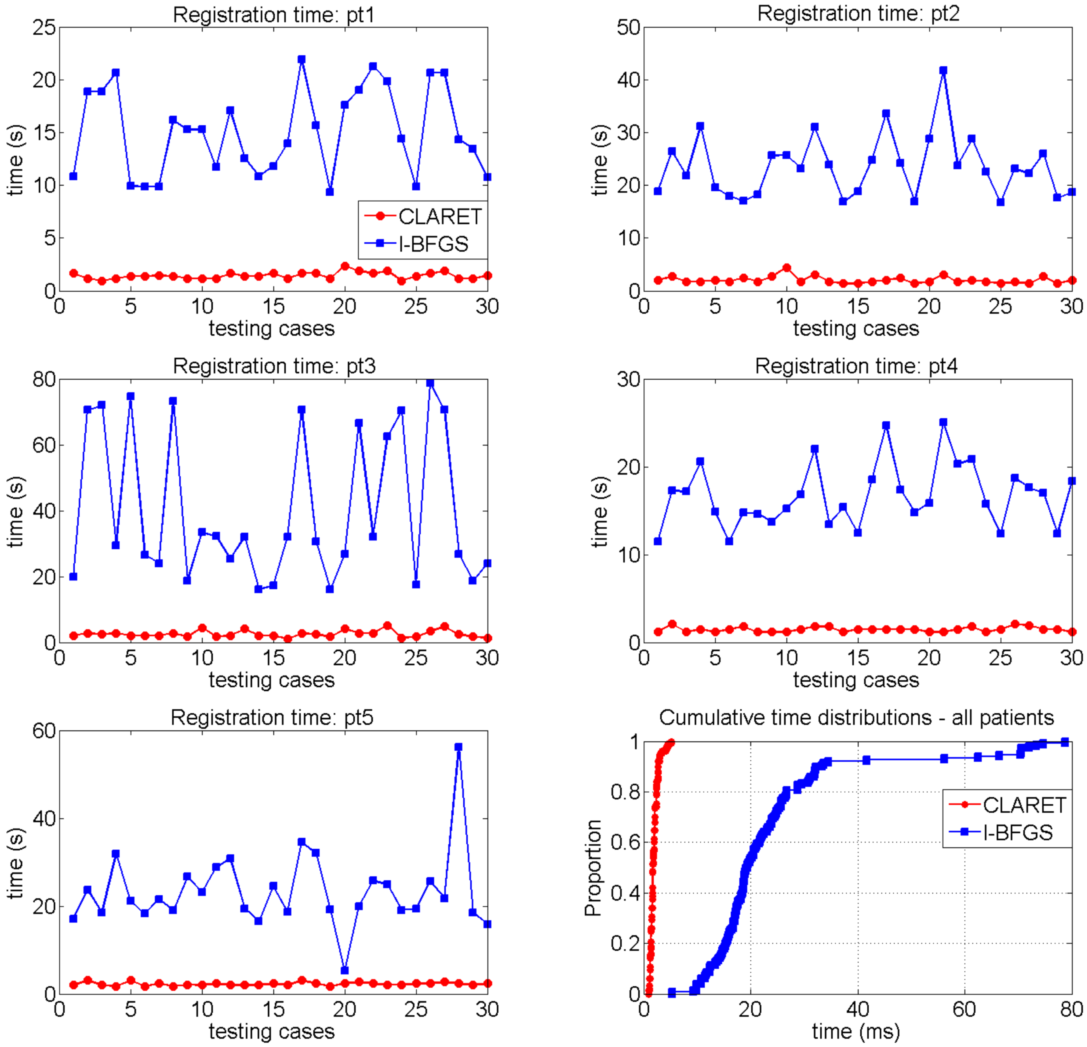

Figure 5.5.9: Registration time on the five patient datasets (pt1-pt5): CLARET vs. the l-BFGS optimization.

Chapter 6: Nonlinear Regression Learning

In this chapter, I present another 2D/3D registration method that relaxes the strong linear assumption made in CLARET. It uses a nonlinear neighborhood analysis approach and calculates the patient's treatment-time 3D deformations by kernel regression. The method is called Registration Efficiency and Accuracy through Learning Metric on Shape (*REALMS*). Specifically, each of the patient's deformation parameters is interpolated using a weighting Gaussian kernel on that parameter's training case values. In each training case, its parameter value is associated with a corresponding training projection image. The Gaussian kernel is formed from distances between training projection images. This distance for the parameter in question involves a Riemannian metric on projection image differences. At planning time, REALMS learns the parameter-specific metrics from the set of training projection images using a Leave-One-Out (*LOO*) training that best fits to the training set.

REALMS applies the metric learning idea first introduced in Weinberger and Tesauro (2007) to the 2D/3D image registration problem. In particular, in order to make the metric learning work for the high dimensional ($D \gg 10^3$) projection space, REALMS uses a specially-designed initialization approximated by CLARET's linear regression as detailed in Section 6.2.2. The results have led to substantial error reduction when the special initialization is applied. As shown in the result, the linear regression implied metrics can directly be used for registration to prevent the time-consuming LOO training.

The rest of the chapter is organized as follows: in Section 6.1, I describe REALMS's novel registration scheme that uses kernel regression. In Section 6.2, I describe the metric learning scheme and the specialized initialization in REALMS. I show synthetic and real results in Section 6.3.

6.1 2D/3D Registration Framework

This section details REALMS's 2D/3D registration framework. REALMS uses kernel regression (Eq. 6.1.1) to interpolate the patient's n 3D deformation parameters $\mathbf{C} = (c^1, c^2, \cdots, c^n)$ separately from the on-board projection image $\Psi(\theta)$ where θ is the projection angle. It uses a Gaussian kernel $K_{\mathbf{M}^i,\sigma^i}$ with width σ^i and a metric tensor $\mathbf{M}^i$ on projection intensity differences to interpolate the patient's i^{th} deformation parameter c^i from a set of N training projection images $\{\mathbf{P}(I \circ T(\mathbf{C}_\kappa); \theta) \mid \kappa = 1, 2, \cdots, N\}$ simulated at planning time. Specifically, the training projection image, $\mathbf{P}(I \circ T(\mathbf{C}_\kappa); \theta)$, is the DRR of a 3D image deformed from the patient's planning-time 3D mean image I with sampled deformation parameters $\mathbf{C}_\kappa = (c^1_\kappa, c^2_\kappa, \cdots, c^n_\kappa)$. T and $\mathbf{P}$ are the warping and the DRR operators, respectively. $\mathbf{P}$ simulates the DRRs according to the treatment-time imaging geometry, e.g., the projection angle θ.

In the treatment-time registration, each deformation parameter c^i in $\mathbf{C}$ can be interpolated with the following kernel regression:

$$c^i = \frac{\sum_{\kappa=1}^{N} c^i_\kappa \cdot K_{\mathbf{M}^i,\sigma^i}(\Psi(\theta), \mathbf{P}(I \circ T(\mathbf{C}_\kappa); \theta))}{\sum_{\kappa=1}^{N} K_{\mathbf{M}^i,\sigma^i}(\Psi(\theta), \mathbf{P}(I \circ T(\mathbf{C}_\kappa); \theta))}, \tag{6.1.1}$$

$$K_{\mathbf{M}^i,\sigma^i}(\Psi(\theta), \mathbf{P}(I \circ T(\mathbf{C}_\kappa); \theta)) = \frac{1}{\sqrt{2\pi}\sigma^i} e^{-\frac{d^2_{\mathbf{M}^i}(\Psi(\theta), \mathbf{P}(I \circ T(\mathbf{C}_\kappa); \theta))}{2(\sigma^i)^2}}, \tag{6.1.2}$$

$$d^2_{\mathbf{M}^i}(\Psi(\theta), \mathbf{P}(I \circ T(\mathbf{C}_\kappa);\theta)) = (\Psi(\theta) - \mathbf{P}(I \circ T(\mathbf{C}_\kappa);\theta))^\intercal \mathbf{M}^i(\Psi(\theta) - \mathbf{P}(I \circ T(\mathbf{C}_\kappa);\theta)), \tag{6.1.3}$$

where $K_{\mathbf{M}^i,\sigma^i}$ is a Gaussian kernel (kernel width= σ^i) that uses a Riemannian metric $\mathbf{M}^i$ in the squared distance $d^2_{\mathbf{M}^i}$ and gives the weights for the parameter interpolation in the regression. The minus signs in Eq. 6.1.3 denote pixel-by-pixel intensity subtraction.

REALMS uses the same deformation space parameterization as CLARET's (See Section 4.2). In the next section, I describe how it learns the metric tensor $\mathbf{M}^i$ and decides the kernel width σ^i.

6.2 Metric Learning at Planning Time

6.2.1 Metric Learning and Kernel Width Selection

REALMS learns a metric tensor $\mathbf{M}^i \subset \mathbb{R}^{\mathcal{P}\times\mathcal{P}}$ with a corresponding kernel width σ^i for the patient's i^{th} deformation parameter c^i using a Leave-One-Out (*LOO*) training strategy. At planning time, it samples a set of N deformation parameter tuples $\{\mathbf{C}_\kappa = (c^1_\kappa, c^2_\kappa, \cdots, c^n_\kappa) \mid \kappa = 1, 2, \cdots N\}$ to generate training projection images $\{\mathbf{P}(I \circ T(\mathbf{C}_\kappa);\theta) \mid \kappa = 1, 2, \cdots, N\}$ where their associated deformation parameters are sampled uniformly within three standard deviations of the scores λ_τ observed in the RCCT. For each deformation parameter c^i in $\mathbf{C}$, REALMS finds the best pair of the metric tensor $\mathbf{M}^{i\dagger}$ and the kernel width $\sigma^{i\dagger}$ that minimizes the sum of squared LOO regression residuals $\mathcal{L}_{c^i}$ among the set of N training projection images:

$$\mathbf{M}^{i\dagger}, \sigma^{i\dagger} = \underset{\mathbf{M}^i,\sigma^i}{arg\,min} \mathcal{L}_{c^i}(\mathbf{M}^i, \sigma^i), \tag{6.2.1}$$

$$\mathcal{L}_{c^i}(\mathbf{M}^i,\sigma^i) = \sum_{\kappa=1}^{N}\left(c^i_\kappa - \hat{c^i_\kappa}(\mathbf{M}^i,\sigma^i)\right)^2, \tag{6.2.2}$$

$$\hat{c^i_\kappa}(\mathbf{M}^i,\sigma^i) = \frac{\sum\limits_{\chi\neq\kappa} c^i_\chi \cdot K_{\mathbf{M}^i,\sigma^i}(\mathbf{P}(I\circ T(\mathbf{C}_\kappa);\theta),\mathbf{P}(I\circ T(\mathbf{C}_\chi);\theta))}{\sum\limits_{\chi\neq\kappa} K_{\mathbf{M}^i,\sigma^i}(\mathbf{P}(I\circ T(\mathbf{C}_\kappa);\theta),\mathbf{P}(I\circ T(\mathbf{C}_\chi);\theta))}, \tag{6.2.3}$$

where $\hat{c^i_\kappa}(\mathbf{M}^i,\sigma^i)$ is the estimated value for parameter c^i_κ interpolated by the metric tensor $\mathbf{M}^i$ and the kernel width σ^i from the training projection images χ other than κ; $\mathbf{M}^i$ needs to be a positive semi-definite ($p.s.d$) matrix to fulfill the pseudo-metric constraint; and the kernel width σ^i needs to be a positive real number.

To avoid high-dimensional optimization over the constrained matrix $\mathbf{M}^i$, I structure the metric tensor $\mathbf{M}^i$ as a rank-1 matrix formed by a basis vector $\mathbf{a}^i \subset \mathbb{R}^{\mathcal{P}\times 1}$:

$$\mathbf{M}^i = \mathbf{a}^i\mathbf{a}^{i\intercal} \tag{6.2.4}$$

Therefore, we can transform Eq. 6.2.1 into an optimization over the unit vector $\mathbf{a}^i$ where $\|\mathbf{a}^i\|_2 = 1$:

$$\mathbf{a}^{i\dagger},\sigma^{i\dagger} = \underset{\mathbf{a}^i,\sigma^i}{argmin}\mathcal{L}_{c^i}(\mathbf{a}^i\mathbf{a}^{i\intercal},\sigma^i) \tag{6.2.5}$$

Then we can rewrite the squared distance $d^2_{\mathbf{M}^i} = d^2_{\mathbf{a}^i\mathbf{a}^{i\intercal}}$ used in the Gaussian kernel $K_{\mathbf{M}^i,\sigma^i}$ as follows:

$$d^2_{\mathbf{a}^i\mathbf{a}^{i\intercal}}(\mathbf{P}(I\circ T(\mathbf{C}_\kappa);\theta),\mathbf{P}(I\circ T(\mathbf{C}_\chi);\theta)) = (\mathbf{a}^{i\intercal}\cdot\mathbf{r}_{\kappa,\chi})^\intercal(\mathbf{a}^{i\intercal}\cdot\mathbf{r}_{\kappa,\chi}), \tag{6.2.6}$$

$$\mathbf{r}_{\kappa,\chi} = \mathbf{P}(I\circ T(\mathbf{C}_\kappa);\theta) - \mathbf{P}(I\circ T(\mathbf{C}_\chi);\theta), \tag{6.2.7}$$

where $\mathbf{r}_{\kappa,\chi} \subset \mathbb{R}^{\mathcal{P}\times 1}$ is a vector of intensity differences between normalized projection images generated by parameters $\mathbf{C}_\kappa$ and $\mathbf{C}_\chi$; and $\mathbf{a}^i$ is a metric basis vector where the magnitude of the inner product of $\mathbf{a}^i$ and the intensity difference vector $\mathbf{r}_{\kappa,\chi}$, $\mathbf{a}^{i\intercal} \cdot \mathbf{r}_{\kappa,\chi}$ gives the distance for the parameter c^i (Eq. 6.2.6).

The learned metric basis vector $\mathbf{a}^{i\dagger}$ and the selected kernel width $\sigma^{i\dagger}$ form a weighting kernel $K_{\mathbf{a}^{i\dagger}\mathbf{a}^{i\dagger\intercal},\sigma^{i\dagger}}$ to interpolate the parameter c^i in the registration (see Eq. 6.1.1).

6.2.2 Linear-Regression Implied Initial Metric

Since the residual functional $\mathcal{L}$ (see Eq. 6.2.1) that we want to minimize is non-convex, a good initial guess of the metric basis vector $\mathbf{a}$ is essential. Therefore, REALMS uses a vector $\mathbf{w}^i$ as an initial guess of the metric basis vector $\mathbf{a}^i$ for the parameter c^i. Let $\mathbf{W} \subset \mathbb{R}^{\mathcal{P}\times n} = \begin{pmatrix} \mathbf{w}^1 & \mathbf{w}^2 & \cdots & \mathbf{w}^n \end{pmatrix}$ list these initial guesses. The matrix $\mathbf{W}$ is approximated by a multivariate linear regression (Eq. 6.2.8 and Eq. 6.2.9) between the projection difference matrix $\mathbf{R} \subset \mathbb{R}^{N\times\mathcal{P}} = (\mathbf{r}_1\mathbf{r}_2\cdots\mathbf{r}_N)^\intercal$ and the parameter differences matrix $\Delta\mathbf{C}$. In particular, the projection difference vector $\mathbf{r}_\kappa = \mathbf{P}(I\circ T(\mathbf{C}_\kappa);\theta) - \mathbf{P}(I;\theta)$ is the intensity differences between the DRRs calculated from the deformed image $I \circ T(\mathbf{C}_\kappa)$ and the DRRs calculated from the mean image I (where $\mathbf{C} = \mathbf{0}$).

$$\Delta\mathbf{C} = \begin{pmatrix} c_1^1 & c_1^2 & \cdots & c_1^n \\ c_2^1 & c_2^2 & \cdots & c_2^n \\ \vdots & \vdots & \ddots & \vdots \\ c_N^1 & c_N^2 & \cdots & c_N^n \end{pmatrix} - \mathbf{0} \approx \begin{pmatrix} \mathbf{r}_1^\intercal \\ \mathbf{r}_2^\intercal \\ \vdots \\ \mathbf{r}_N^\intercal \end{pmatrix} \cdot \begin{pmatrix} \mathbf{w}^1 & \mathbf{w}^2 & \cdots & \mathbf{w}^n \end{pmatrix} = \mathbf{W}^\intercal\mathbf{R} \quad (6.2.8)$$

$$\mathbf{W} = (\mathbf{R}^\intercal\mathbf{R})^{-1}\mathbf{R}^\intercal\Delta\mathbf{C} \quad (6.2.9)$$

This is just the regression that was used to train CLARET.

The inner product of the matrix $\mathbf{W}$, calculated by the pseudo-inverse in Eq. 6.2.9, and the projection intensity difference matrix $\mathbf{R}$, $\mathbf{W}^{\intercal}\mathbf{R}$, gives the best linear approximation of the parameter differences $\Delta\mathbf{C}$ from the mean image ($\mathbf{C} = \mathbf{0}$). Therefore, we use $\mathbf{w}^i$ as the initial guess of the metric basis vector $\mathbf{a}^i$ for the parameter c^i.

As alternative to using this LOO process of optimization, this linear regression implied metric can be directly used as the final metric for parameter interpolation at registration time. This allows an efficient metric learning that prevents the time-consuming LOO training.

6.2.3 Optimization Scheme

REALMS uses a two-step scheme to optimize the metric basis vector $\mathbf{a}^i$ and the kernel width σ^i in Eq. 6.2.5.

First, for each candidate kernel width σ^i, it optimizes the metric basis vector $\mathbf{a}^i$ using the quasi-Newton method (specifically, the BFGS method) with the vector $\mathbf{w}^i$ as the initialization. The gradient of the function $\mathcal{L}_{c^i}$ with respect to $\mathbf{a}^i$ can be stated as

$$\frac{\partial \mathcal{L}_{c^i}}{\partial \mathbf{a}^i} = \frac{2\sqrt{2}}{\sigma^i}\mathbf{a}^i \sum_{\kappa=1}^{N}(\hat{c^i_\kappa} - c^i_\kappa) \sum_{\chi=1}^{N}(\hat{c^i_\chi} - c^i_\chi) K_{\mathbf{a}^i\mathbf{a}^{i\intercal},\sigma^i}(\mathbf{P}(I \circ T(\mathbf{C}_\kappa);\theta), \mathbf{P}(I \circ T(\mathbf{C}_\chi);\theta))\mathbf{r}_{\kappa,\chi}\mathbf{r}^{\intercal}_{\kappa,\chi} \tag{6.2.10}$$

Second, REALMS selects a kernel width $\sigma^{i\dagger}$ among the candidate kernel widths where its learned metric basis vector $\mathbf{a}^{i\dagger}$ yields minimum LOO regression residuals $\mathcal{L}_{c^i}$ for parameter c^i.

6.3 Results

REALMS's registration accuracy and efficiency were tested by synthetic (6.3.1) and real (6.3.2) datasets. In Chapter 8, I also compare its accuracy and efficiency to CLARET and the state-of-the-art optimization-based method (Li *et al.* (2011a)).

6.3.1 Synthetic Tests

Coronal DRRs (dimension: 64×48) of the target CTs were used as synthetic on-board cone-beam projection images. The target CTs were deformed from the patient's Fréchet mean CT by normally distributed random samples of the first three deformation parameters.[1] 600 synthetic test cases were generated from 6 lung datasets and the registration quality was measured by the average *mTRE* (mean Target Registration Error) over all cases and all voxels at tumor sites. With REALMS's registrations, the average mTRE and its standard deviation are down from 6.89 ± 3.53 mm to 0.34 ± 0.24 mm using $N = 125$ training projection images. The computation time for each registration is 11.39 ± 0.73 ms (87.79 registrations per second) on an Intel Core2 Quad CPU Q6700.

As shown in Fig. 6.3.1, REALMS reduces the minimum errors produced by kernel regressions that use the Euclidean metric ($\mathbf{M}^i = \mathbf{I}$). In particular, the REALMS metric learned by the LOO optimization best minimizes the error for all three parameters. However, per results shown in the figure, the REALMS's initial metric learned by linear regression (CLARET) already yielded good registration accuracy.

---

[1] In my lung datasets, the first three deformation parameters captured more than 95% lung variation observed in their RCCT.

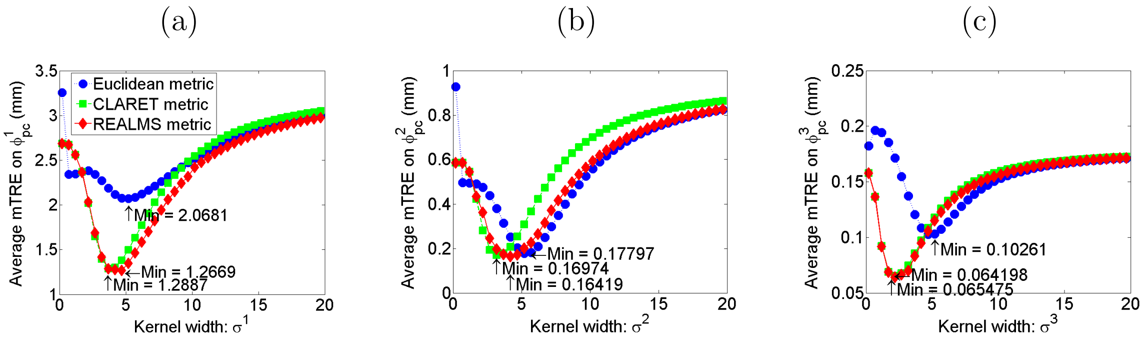

Figure 6.3.1: Average mTREs over 600 test cases projected onto the (a) first, (b) second, and (c) third deformation basis vector versus the candidate kernel widths using $N = 125$ training projection images.

Fig. 6.3.2 shows the computation time and registration accuracy tradeoff in REALMS.

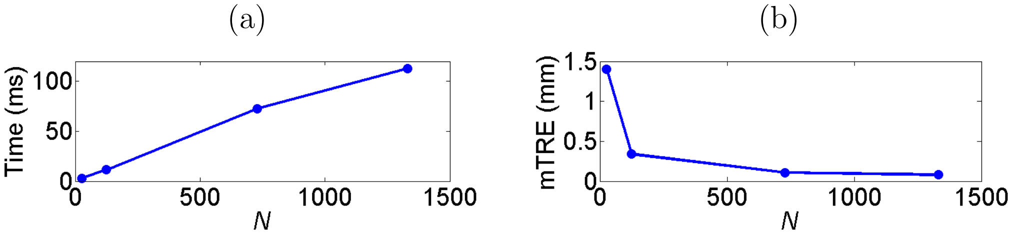

Figure 6.3.2: (a) Time and (b) accuracy v.s. the number of training projection images N.

6.3.2 Real Tests

REALMS was also tested on 6 lung datasets with an on-board CBCT system where a *single* coronal on-board CB projection (dimension downsampled to 64×48 for efficient computation) at both *EE* (End-Expiration) and *EI* (End-Inspiration) phases were used for the testing. See the top image of Fig. 6.3.4(b) for illustration. For each dataset, we generated $N = 125$ training DRRs to learn the metrics and select optimal interpolation kernel widths. The learned metrics and the selected kernel widths were used to estimate deformation parameters for the testing of EE and EI on-board projections. The estimated CTs were deformed from the Fréchet mean CT with the estimated deformation parameters. The results were validated with reconstructed CBCTs at target phases.[2] Table 6.1 shows the 3D Tumor Centroid Differences (*TCD*s) between REALMS-estimated CTs and the reconstructed CBCTs at the same respiratory phases. Tumor centroids were computed via Snake active segmentations. As shown in Table 6.1, REALMS reduces the TCD from 5.58 ± 3.14 mm to 2.56 ± 1.11 mm in 10.89 ± 0.26 ms (91.82 fps).

dataset#	TCD at EE phase (mm)	TCD at EI phase (mm)	Time (ms)
1	2.42 (9.70)	4.06 (7.45)	10.40
2	3.60 (4.85)	3.60 (4.89)	10.92
3	2.30 (8.71)	3.60 (4.03)	10.91
4	1.27 (2.69)	2.80 (2.29)	10.91
5	0.70 (9.89)	3.28 (8.71)	11.15
6	1.98 (2.03)	1.12 (1.72)	11.08

Table 6.1: Tumor Centroid Differences (*TCD*) after REALMS's registration at EE and EI phases of 6 lung datasets. Numbers inside the parentheses are the initial TCDs.

Fig. 6.3.3 illustrates an example REALMS registration on a lung dataset where the tumor, the diaphragm, and most of the soft tissues are correctly aligned.

---

[2] The CBCTs were reconstructed by the retrospectively-sorted CB projections at target breathing phases.

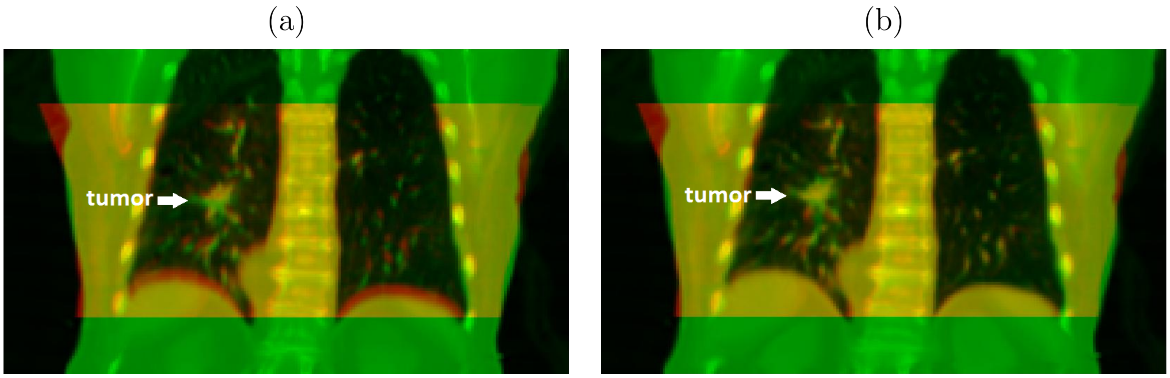

Figure 6.3.3: (a) Image overlay of the reconstructed CBCT at EE phase (red) and the Fréchet mean CT (green) (b) Image overlay of the reconstructed CBCT at EE phase (red) and the REALMS-estimated CT (green) calculated from an on-board cone-beam projection image at EE phase. The yellow areas are the overlapped region.

6.3.3 The Learned Metric Basis Vector

The learned metric basis vector $\mathbf{a}^{i\dagger}$ will emphasize projection pixels that are significant for the distance calculation of the deformation parameter c^i (e.g., give high positive or high negative values). As shown in Fig. 6.3.4(a), the learned metric basis vector $\mathbf{a}^{1\dagger}$ emphasized the diaphragm locations and the lung boundaries as its corresponding deformation basis vector ϕ_{pc}^1 covers the expansion and contraction motion of the lung. See the bottom image of Fig. 6.3.4(b) for illustration.

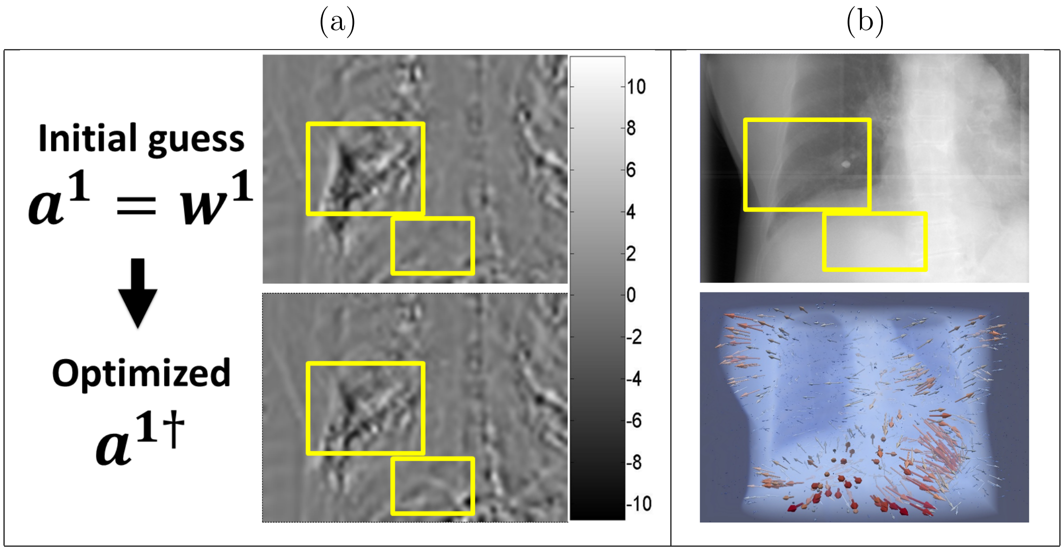

Figure 6.3.4: (a) Initial guess of the metric basis vector $\mathbf{a}^1 = \mathbf{w}^1$ (top) and the optimized metric basis vector $\mathbf{a}^{1\dagger}$ (bottom) of a lung dataset. They are re-shaped into the projection image domain for visualization. Both very dark and very light pixels contribute the most to the distance calculation of the parameters. As shown in the figure, the diaphragm locations and the lung boundaries (yellow boxes) were emphasized after metric learning. (b) Top: a coronal on-board CB projection at EE phase of the lung dataset used in (a). The yellow boxes in (a) and (b) correspond to the same 2D locations. Bottom: the first deformation basis vector ϕ^1_{pc} (the color arrows indicate heated object spectrum maps of the deformation magnitudes) overlaid with the volume rendering of the Fréchet mean CT of the lung dataset used in (a). For this dataset, ϕ^1_{pc} covers the expansion and contraction motion of of the lung.

Chapter 7: Locally-linear Regression Learning

As presented in Ch. 5, CLARET learns global linear regressions that map projection intensities to their associated deformation parameters based on a sampling from the whole deformation space. At treatment time the learned patient-specific regressions are iteratively applied to refine the estimation of the patient's deformation parameters. In Ch. 6, I showed that REALMS is an even faster regression learning method that does not need to iterate for refinement. It estimates the patient's deformation parameters by learning good global projection distance metrics for deformation parameter interpolation via nonlinear kernel regressions.

However, due to the variable projection-to-deformation relationships in various regions of the deformation space, a global regression learned from the whole deformation space is a rough approximation of the underlying relationship. As a result, the registration accuracy is limited by the global learning methods.

This chapter presents a novel local learning method that partitions the deformation space, learns a projection-to-deformation relationship for each partition and at application time identifies the nearest partition for an unseen x-ray image and applies the learned local relationship of the identified partition to yield better approximation. The method improves the regression learning method described in Ch. 5 as an example. The results show that the method's local regression learning yields a more accurate and still efficient 2D/3D deformable registration.

7.1 Method Overview

The method generates large-scale training samples and finds normalized graph cuts (Shi and Malik (2000)) that best separate the training samples into a given number of training partitions. In each training partition the method learns linear regression matrices that map the training projection intensities to the training deformation parameters. At treatment time, the method decides which training partition the target deformation resides in by a trained decision forest (Criminisi *et al.* (2011), an approach that has shown success in many medical applications) based on projection image intensities and gradients of various image regions in the target projection image. The linear regressions learned for the forest-decided partition are applied to the target projection intensities to yield the desired deformation.

The purpose of the local regression learning is to obtain a better regression fitting to the training set. Due to the nonlinear relationships between the deformation parameters and the projection intensities, fitting a globally linear regression to approximate this nonlinear relationship is mathematically inappropriate.

This section presents a locally-linear regression learning method that can more flexibly approximate this nonlinear relationship. It comprises two stages: training and treatment application. There are five steps in the training stage. First, the patient's prior deformation space is parameterized by an LDDMM (Large Deformation Diffeomorphic Metric Mapping) framework from the patient's treatment-planning RCCT. Second, the method samples training deformation parameters and simulates corresponding x-ray projection images, or DRRs (Digitally-Reconstructed Radiographs), from CT volumes warped by the sampled training deformations. Third, the method partitions the training projection images recursively by normalized graph cuts using Euclidean distances of the deformation parameters. Fourth, the method computes a linear regression between the deformation parameters and the covarying projection intensities for each training partition. Finally, in order to classify an unseen projection image into a nearest training partition in the treatment application stage, in the training stage the method trains a set of decision trees, or a decision forest, for deciding the local training partition based on the projection image intensities and gradients of various image regions. In the treatment application stage, given a target projection image the method estimates the deformation parameters of the patient by first classifying the target projection image into a training partition and then using the learned linear regression of that partition to yield the estimation.

7.2 Training Stage

7.2.1 Deformation Space Formulation

The method uses the deformation space formulation described in Ch. 4. Recall that the method constrains the patient's deformation to a space spanned from the deformation observed in the patient's treatment-planning RCCT consisting of 10 phase images. The weights λ_τ^i on each deformation eigenmode i yield a deformation ϕ_τ in terms of these deformation eigenmodes: $\phi_\tau = \overline{\phi} + \sum_{i=1}^{10} \lambda_\tau^i \cdot \phi_{pc}^i$.

7.2.2 Training Space Sampling

In preparation of the partitioning, the method uniformly samples S scores on each deformation eigenmode within ±3 standard deviations of the eigenmode scores observed in the planning RCCT. For 3-dimensional parametrization of the deformation space, a total $N = S^3$ sampled deformation parameters $\{\mathbf{C}_\kappa\}_{\kappa=1,2,\ldots,N}$ are used to warp the Fréchet mean image, and the training projection images $\{\mathbf{P}_\kappa\}_{\kappa=1,2,\ldots,N}$ are simulated from those warped mean images. In order to make the intensity comparable between the training projections and treatment-time target projections, the method normalizes both training and target projections with a local Gaussian normalization method and a histogram matching method described in Sections 3.4.1 and 3.4.2.

7.2.3 Training Space Partitioning

To partition the training projections, the method uses a hierarchical separation approach that is similar to the normalized graph cuts method (Shi and Malik (2000)). The graph cuts are decided based on the Euclidean distance of the deformation parameters, a criterion similar to that used in the multidimensional scaling (MDS) method (Kruskal (1964)).

The goal of the training space partitioning is to separate the training space such that the Euclidean distance of the deformation parameters in each partition is minimized. Different from the traditional graph cut approach that generates an "affinity" matrix, the method generates an $N \times N$ "dissimilarity" matrix $\mathcal{D}$ where the entries $\mathcal{D}_{i,j} = \|\mathbf{C}_i - \mathbf{C}_j\|_2$ are the distances of the deformation eigenmode weights between training samples i and j. N is the number of training samples. Having the dissimilarity matrix computed, like previous literature stated in (Shi and Malik (2000)) the method computes the first k smallest eigenvalues and their eigenvectors $V_1, V_2, .., V_k$ of the dissimilarity matrix $\mathcal{D}$ if 2^k partitions are needed. The eigenvector with the smallest eigenvalue, V_1, is an approximation to the NP-hard normalized graph cut problem: training samples close to each other will have similar values in the eigenvector (have close dissimilarity maps). Therefore, the method sorts the training samples by their values in V_1 and partitions the training samples into two by the median value. The final partitioning is computed by recursively separating the training set using the eigenvectors with the next smallest eigenvalues.

7.2.4 Local Regression Learning

The method approximates the nonlinear relationship between the projection intensities and the deformation parameters by fitting a linear regression $\mathbf{W}^i_{\mathcal{L}}$ that linearly maps the projection intensities to the i^{th} deformation parameter $c^i_{\kappa\in\mathcal{L}}$ for each training sample κ in each local training partition $\mathcal{L}$ (local deformation neighborhood): $c^i_{\kappa\in\mathcal{L}} \approx \mathbf{P}_{\kappa\in\mathcal{L}} \cdot \mathbf{W}^i_{\mathcal{L}}$ where $\mathbf{W}^i_{\mathcal{L}} = (\mathbf{P}^{\mathsf{T}}_{\kappa\in\mathcal{L}} \cdot \mathbf{P}_{\kappa\in\mathcal{L}})^{-1} \cdot \mathbf{P}^{\mathsf{T}}_{\kappa\in\mathcal{L}} \cdot c^i_{\kappa\in\mathcal{L}}$.

7.2.5 Decision Forest Training

In order to efficiently classify an unseen target projection into a training partition in the application stage, the method constructs a decision forest $\mathcal{F}$ in the training stage. The decision forest $\mathcal{F}$ consists of M binary trees $\mathcal{T}_1, \mathcal{T}_2, .., \mathcal{T}_M$ with depth d. In this work $M = 100$ and $d = 5$ are used. In each tree, a tree traversal of the target projection selects the nearest training projection images. With the 100 tree traversals the method classifies the target projection image into the same partition as its most frequently selected nearest training projection image's. The tree traversal is guided by a sequence of binary decisions made at tree nodes based on the "visual features" of the target projection image. The visual features used in this work are 6: mean intensity, mean intensity difference, mean intensity gradients along two projection axes, and mean intensity gradient differences along two projection axes of $1,000$ random box pairs (random positions and sizes).

The method constructs such a decision forest through a supervised learning on the training set: in each tree the binary decisions made at tree nodes are to select visual features and their thresholds to partition the training projection images such that training projection images with close deformation parameter values will traverse to the same leaf node. Specifically, for each binary tree $\mathcal{T}_\delta$ the method randomly selects 400 candidate visual features $\mathcal{V}_\delta$ out of the total $1,000 \times 6 = 6,000$ visual features $\mathcal{V}$. At each tree node $\mathcal{N}$ in the tree $\mathcal{T}_\delta$ the method selects the q^{th} visual feature $\mathcal{V}_\mathcal{N}^q$ among the 150 candidate visual features $\mathcal{V}_\mathcal{N}$ randomly sampled from $\mathcal{V}_\delta$ ($\mathcal{V}_\mathcal{N}^q \in \mathcal{V}_\mathcal{N} \subset \mathcal{V}_\delta \subset \mathcal{V}$) and the threshold ξ for the selected visual feature $\mathcal{V}_\mathcal{N}^q$ such that after the binary partition by the threshold, the total variance of the deformation parameters is minimized or equivalently, the information gain is maximized: for training samples κ in tree node $\mathcal{N}$,

$$\underset{q,\xi}{\mathbf{argmax}}\ \mathbf{Var}\{\mathbf{C}_{\kappa\in\mathcal{N}}\} - \mathbf{Var}\{\mathbf{C}_{\kappa\in\mathcal{N}} | \mathcal{V}_\mathcal{N}^q(\mathbf{P}_\kappa) \geqslant \xi\} - \mathbf{Var}\{\mathbf{C}_{\kappa\in\mathcal{N}} | \mathcal{V}_\mathcal{N}^q(\mathbf{P}_\kappa) < \xi\} \quad (7.2.1)$$

With optimizations (7.2.1) at all tree nodes in all trees, training projection images traversing to the same leaf nodes in a tree have the same visual feature responses to the binary decisions and similar deformation parameter values. The random selection of the candidate features for each tree and for each node provides an efficient and robust discriminative learning from the high dimensional feature space (Breiman (2001)). The method records the leaf node indices that the training projection images visited in each tree for classification in the treatment application stage.

7.3 Treatment Application Stage

In the treatment application stage, a previously unseen target projection image is preprocessed to remove the additional photon scattering using the local Gaussian normalization method and histogram matching method described in Sections 3.4.1 and 3.4.2. The method then classifies the target projection image into a local training partition by the trained regression forest. Finally, the target deformation parameters are estimated by the local linear regressions learned from that local partition. With the efficient forest classification and the fast parameter estimation by regressions that only involve matrix multiplication, the whole registration process can be computed in real time.

7.3.1 Forest Classification

The classification consists of two steps: first, each tree traversal of the target projection image yields a neighborhood set of training projection images that visit the same leaf node. For 100 trees the method computes the most frequently visited training projection image from the 100 neighborhood sets. Second, the method assigns the target projection image the same partition as its nearest training neighbor's.

7.3.2 Regression Estimation

With the training partition $\mathcal{L}^*$ classified, the method uses the learned linear regression $\mathbf{W}^i_{\mathcal{L}^*}$ of that training partition to estimate the deformation parameter c^i_{target} from the target projection intensities $\mathbf{P}_{target}$: $c^i_{target} = \mathbf{P}_{target} \cdot \mathbf{W}^i_{\mathcal{L}^*}$.

7.4 Results

The local regression method has been validated on 500 synthetic treatment-time deformations and 5 real treatment-time deformations sampled from 5 lung datasets (pt1 to pt5). In the synthetic tests, target projection images were the DRRs computed from the Fréchet mean image warped by random treatment-time deformations sampled from the patient's deformation eigenmode space within ±3 standard deviations of the weights observed in the patient's RCCT. Having the ground truths of the deformation parameters, the synthetic tests were used to analyze the optimal training settings (e.g., the number of training samples and partitions) in terms of the method's registration accuracy and efficiency. Moreover, to better understand the importance of the local regression, the accuracy of the forest classification and the accuracy of the local regression are also compared in this section. In the next chapter (Chapter 8), I also compare its accuracy and efficiency to CLARET, REALMS, and the state-of-the-art optimization-based method (Li *et al.* (2011a)).

7.4.1 The Datasets

Five lung datasets have been tested for this local regression method. Each dataset consists of 10 treatment-planning RCCT, a target cone-beam CT (CBCT) projection (coronal-view, dimension down-sampled to 128×96) scanned at the patient's end-of-expiration (EE) phase, and a validating 3D EE-phase CBCT reconstructed at treatment time. As shown in Fig. 7.4.1 those five datasets represent very different pathological states and imaging fields of views (FOVs). For example, patient 1 and 4 (pt 1 and 4) have a shorter CT scan along the superior-inferior (SI) direction and therefore the simulated projections have truncations; Patient 4 has only one lung, and patient 5 has an extended view toward the abdominal region. For all five patient datasets, their target projections were all imaged in the patients' EE phases. The estimated deformations are validated by treatment-time reconstructed cone-beam CTs (CBCTs) at the same EE phases. In particular, the results measure mean target registration errors (mTREs) as (1) the tumor centroid differences (for real tests) or (2) the average deformation differences of a lung voxel (for synthetic tests) between the estimated CTs (Fréchet mean images warped by the estimated deformations) and the validating CBCTs (for real tests) or the ground truth target CTs (for synthetic tests). Tumors were manually segmented in both CTs and CBCTs.

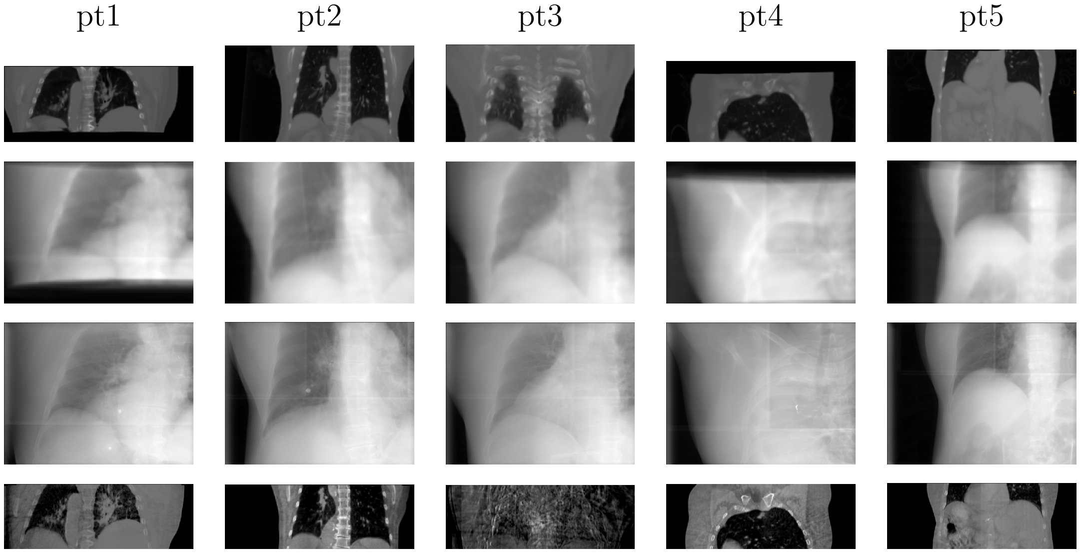

Figure 7.4.1: The 5 lung datasets. Top row: the middle coronal slices of the patients' Fréchet mean CT images. Second row: DRRs of the Fréchet mean CT images. Third row: target projection images at the end-of-expiration (EE) phase. Fourth row: validating reconstructed cone-beam CTs (CBCTs) at the EE phases.

7.4.2 Synthetic Tests

For each lung dataset, 100 testing projections and testing deformation parameters were randomly sampled in the patient's deformation eigenmode space. The registration accuracy and efficiency have been tested with a varying number of training samples and with varying number of partitions.

Training Space Sampling and Partitioning

Five different numbers of partitions have been chosen for testing: 1, 2, 4, 8, and 16 partitions. The method partitions the training samples recursively by normalized graph cuts. As shown in Fig. 7.4.2, the method first separates the training space along the first deformation eigenmode because it contributes the greatest variation of the training deformations.

Forest Classification

Also shown in Fig. 7.4.2, the decision forest successfully classifies most of the testing samples into the correct training partitions. The erroneous partition assignments only happen on the partition boundaries.

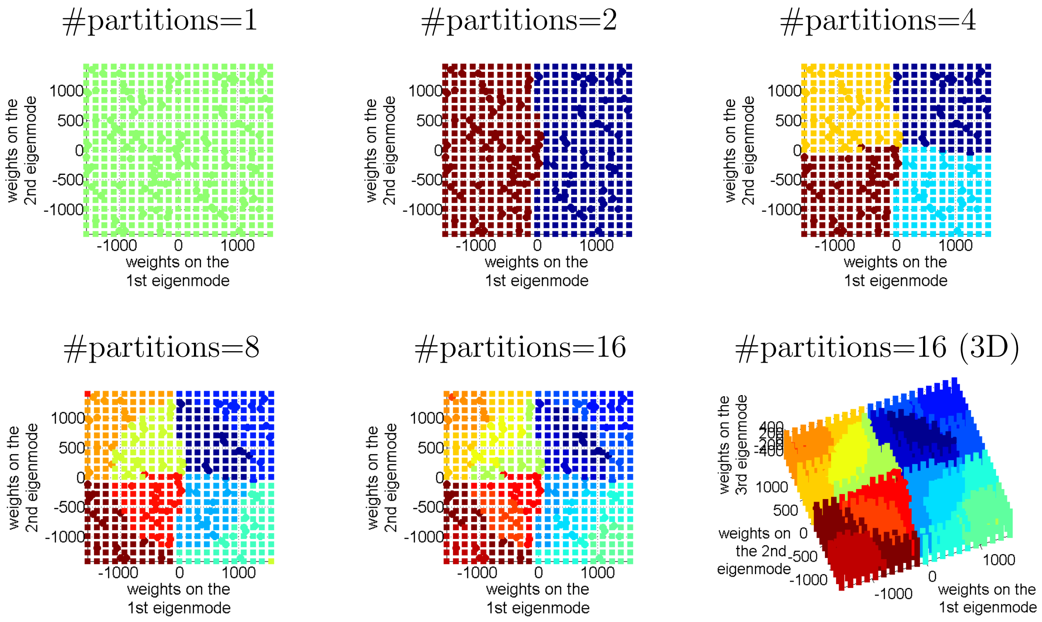

Figure 7.4.2: Training samples partitioning and forest classification results of a lung dataset. Circular dots are 21^3 training samples in the deformation eigenmode space. Square dots are the 100 random testing samples colored by different partitions decided by the forest.

Forest Classification vs. Local Regression Accuracy

Forest classification selects a nearest training neighbor for each testing target projection image. The deformation parameters of the selected nearest neighbor can be used as a rough estimation, and the further local regression provides refinements of the forest classification. To demonstrate the refinement improvement of the local regression, 500 synthetic tests were generated from the 5 lung datasets (100 tests each). The accuracy is measured by the mean deformation error appearing in the lung. As shown in Fig. 7.4.3, both forest classification and the local regression accuracy improved greatly with an increasing number of training samples. However, the accuracy of the forest classification converges to 0.5 mm whereas the local regression converges to 0.05 mm, which shows a 10-fold error reduction.

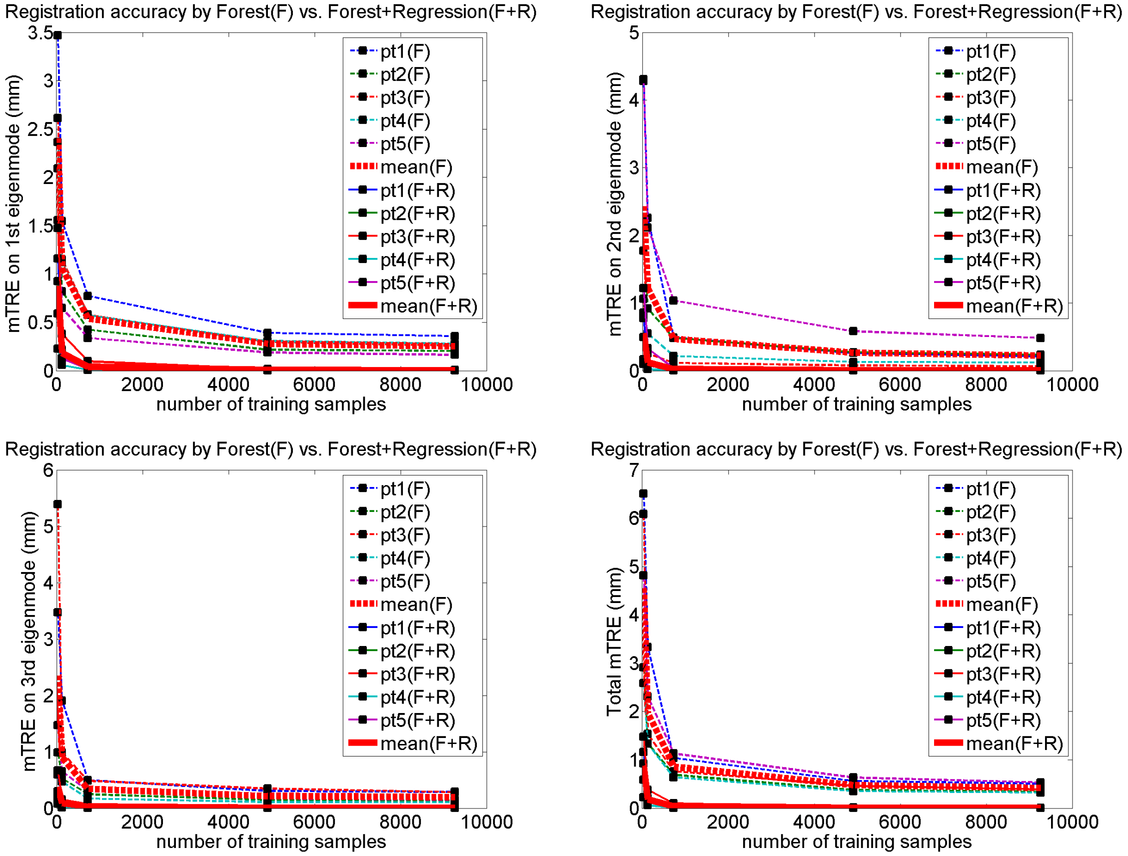
Registration accuracy by Forest(F) vs. Forest+Regression(F+R)
Registration accuracy by Forest(F) vs. Forest+Regression(F+R)
Registration accuracy by Forest(F) vs. Forest+Regression(F+R)
Registration accuracy by Forest(F) vs. Forest+Regression(F+R)

Figure 7.4.3: Average deformation errors of a lung voxel in the 100 synthetic tests for each patient with varying numbers of training samples. The registration accuracy of the forest classification (F) is compared with the accuracy of the forest classification followed by local regression (F+R). In the figures the local regression uses the optimal number of training partitions that yields the most accurate registration.

The More Partitions, the Better?

Registration time and accuracy have been investigated with the number of the training partitions used. Fig. 7.4.4 shows the registration error slightly reduces and then increases with the number of training partitions. This suggests that the local regression enhances the registration accuracy if the a proper partition of the training space is applied. For the results shown in the figure, with the $N = 21^3$ training samples the linear regressions best fit to a smaller partition rather than to the entire training set. The increasing errors observed in the over-partitioning situations (the number of partitions > 8) are the results of overfitting to partitions of insufficient training samples.

In terms of registration time, since the global regression learning does not require the forest classification, the mean computation time for the regression inference is only 10 ms. For local regression learning, Fig. 7.4.5 shows that it yields minimum registration time (forest classification time + local regression time) when 4 training partitions are used, and the registration is still real-time.

As per the more accurate registration accuracy and the still-in-real-time speed, the local regression learning is better than the global regression learning.

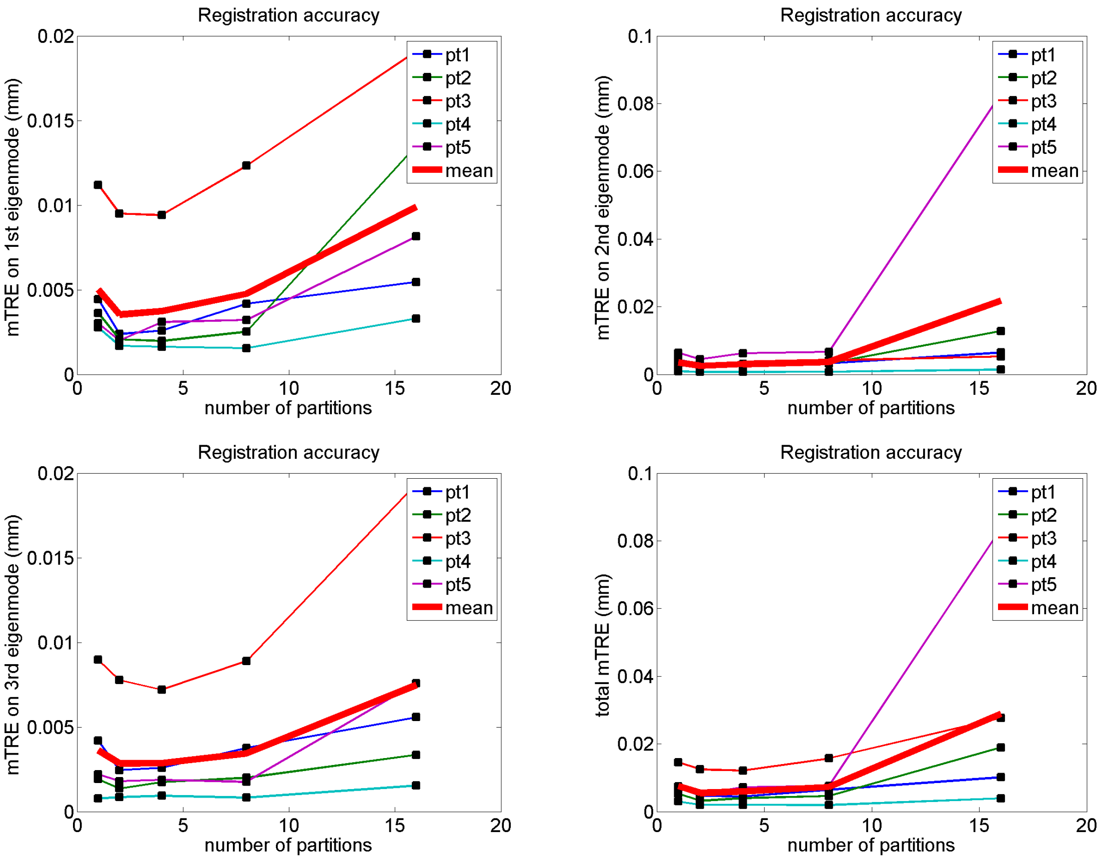

Figure 7.4.4: Average deformation errors of a lung voxel in the 100 synthetic tests for each patient with varying numbers of training partitions for 21^3 training samples.

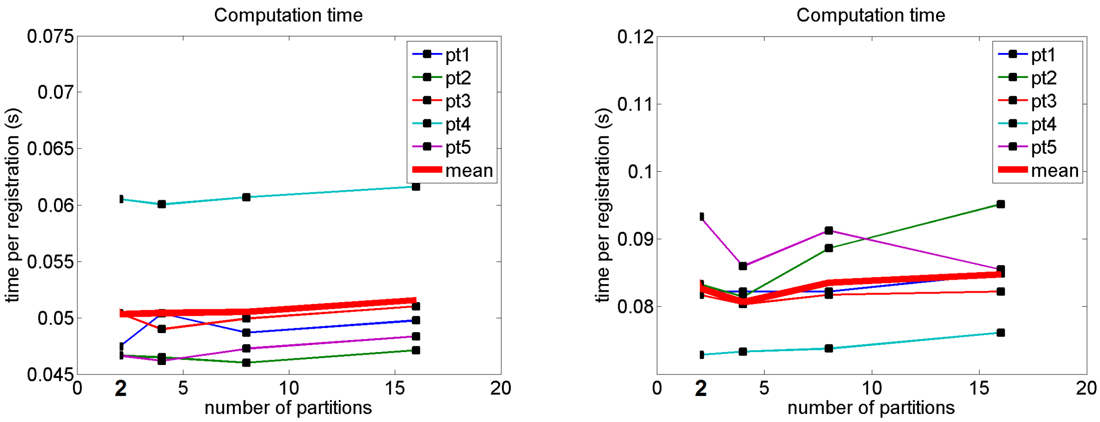

Figure 7.4.5: Average registration time of the 100 synthetic tests for each patient with varying numbers of training partitions for 17^3 (left) and 21^3 (right) training samples. Optimal registration time is more distinguishable when using larger training sample sets.

7.4.3 Real Tests

Results on Real Datasets by the Optimal Partitioning Learned from Patient-Specific Synthetic Tests

For registration of the real datasets, Table 7.1 compares the registration time and accuracy between the global and local learning methods. Results of the local learning method use the optimal training settings suggested by the synthetic tests, e.g., for patient 2, the number of training samples=21^3 and the number of partitions=4. As shown in the table, the local learning method is more accurate than the global learning method. Its computation time is still real-time even with the extra forest classification (that is not used by the global learning). The mean 3D tumor centroid errors e^{3D} over the five datasets are 3.73 mm and 2.61 mm for the global and the local learning, respectively. The error reduction made by this local learning has no particular trend in favoring specific directions (neither directions in-plane nor the out-of-plane direction). In addition, for target projection images having a more abdominal FOV like patient

5's, the great error reduction (from 6.65 to 2.78 mm) shows the method's ability to learn the nonlinear relationship between the projection intensities and the deformation parameters as the abdomen undergoes more nonlinear motions than the lung.

Global learning	pt1	pt2	pt3	pt4	pt5
#partitions	1	1	1	1	1
e^{3D} (mm)	4.18	1.94	3.78	2.12	6.65
e^{2D} (mm)	4.16	1.84	3.74	1.88	6.52
$e^{\perp}$(mm)	0.4	0.6	0.54	0.97	1.27
Time (s)	0.001	0.001	0.001	0.001	0.001

Local learning	pt1	pt2	pt3	pt4	pt5
#partitions	4	4	4	4	2
e^{3D} (mm)	3.23(22%)	1.85(4.6%)	3.08(18%)	2.11(0.4%)	2.78(58%)
e^{2D} (mm)	3.19(23%)	1.78(-0.5%)	3.03(19%)	1.97(-4.7%)	2.70(58%)
$e^{\perp}$(mm)	0.5(-24%)	0.5(16%)	0.52(3.7%)	0.74(23%)	0.64(49%)
Time (s)	0.06	0.07	0.07	0.06	0.07

Table 7.1: Registration time and accuracy of the five real datasets. The accuracy is measured by the tumor centroid differences, e, between the estimated CTs and the validating CBCTs. e^{3D}: 3D errors; e^{2D}: errors on the patient plane (coronal plane); $e^{\perp}$: errors orthogonal to the patient plane (out-of-plane direction). Numbers inside the parentheses are the percentage error reduction from global learning.

Is the Learned Optimal Partitioning Still Optimal for the Real Dataset?

Fig. 7.4.6 shows the registration accuracy and time vs. the number of training partitions of the 5 real datasets. As shown in the figure, the optimal number of training partitions that yields the smallest registration error is 4, which agrees with the synthetic results, except that the advantage of having 4 partitions is stronger here than with the synthetic tests. However, different from the synthetic results, the registration error does not increase when more than 4 training partitions are used. One possible cause for this inconsistency is that the optimal number of training partitions depends on the distribution of the target projection images in the training deformation space. For example, the synthetic target projection images are distributed in the whole training deformation space whereas the real target projection images are all in the EE phases, which lie in a region close to the boundaries of the training deformation space (see boundaries in the sub-figures in Fig. 7.4.2). Consequently, regressions fitted to a smaller training partition will tend to have better estimation of this extreme testing deformation.

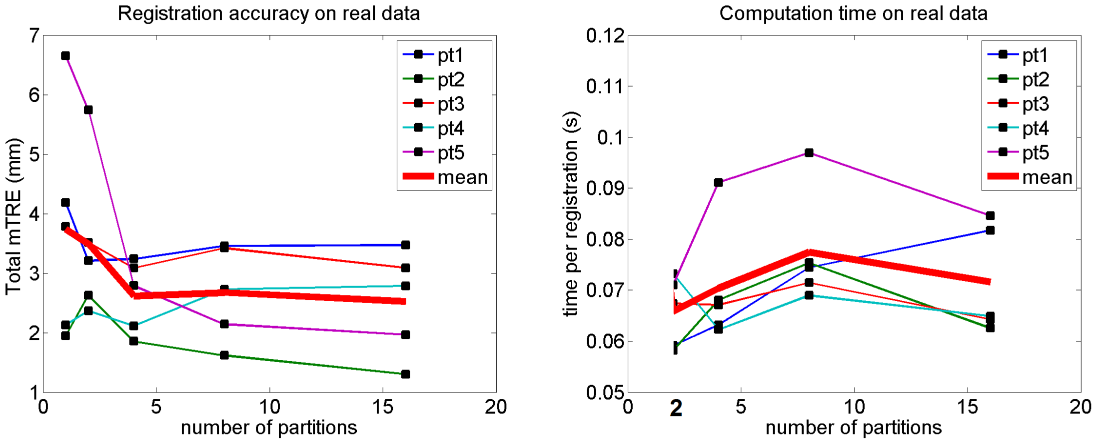

Figure 7.4.6: Registration accuracy (left) and time (right) vs. the number of training partitions on 5 real lung datasets.

Chapter 8: Comparisons

Registration accuracy has also been compared between the proposed regression-based methods and the optimization-based method. Section 8.1 compares registration results on synthetic data obtained from the patient-specific RCCT datasets, and Section 8.2 compares registration results on real CB projection data. Discussion on picking the winning method among the proposed methods is presented in Section 8.3.

8.1 Synthetic Tests

For each method Fig. 8.1.1 shows the cumulative error distribution of a total of 150 synthetic random tests generated from the 5 patient datasets. Initial errors ("init" in the legend of the figure) in the 150 test cases range from 1 to 10 mm. The registration errors are reported as the mean Target Registration Error (mTRE) of a lung voxel over millions of voxels over the 150 synthetic random tests. Since the deformation is zero in almost all the voxels, this yields sub-millimeter mTREs even though the maximum error is 10 mm. However, for comparisons this error measurements are still valid. Detailed explanations of the acronyms in Fig. 8.1.1 are as follows:

- init: initial deviations obtained from the mean CT.
- l-BFGS: The state-of-the-art optimization-based registration using a quasi-Newton l-BFGS optimizer.
- CLARET(125)-1(itr.): iterative CLARET using 125 training images.

- CLARET(125)-1: noniterative global CLARET using 125 training images.
- Forest(125): forest classification using 125 training images.
- REALMS(125)-1: REALMS registration using 125 training images.
- CLARET(9261)-1: noniterative global CLARET using 9261 training images.
- CLARET(9261)-2: noniterative local CLARET using 9261 training images and 2 partitions.
- CLARET(9261)-4: noniterative local CLARET using 9261 training images and 4 partitions.
- CLARET(9261)-8: noniterative local CLARET using 9261 training images and 8 partitions.
- CLARET(9261)-16: noniterative local CLARET using 9261 training images and 16 partitions.
- Forest(9261): forest classification using 9261 training images.

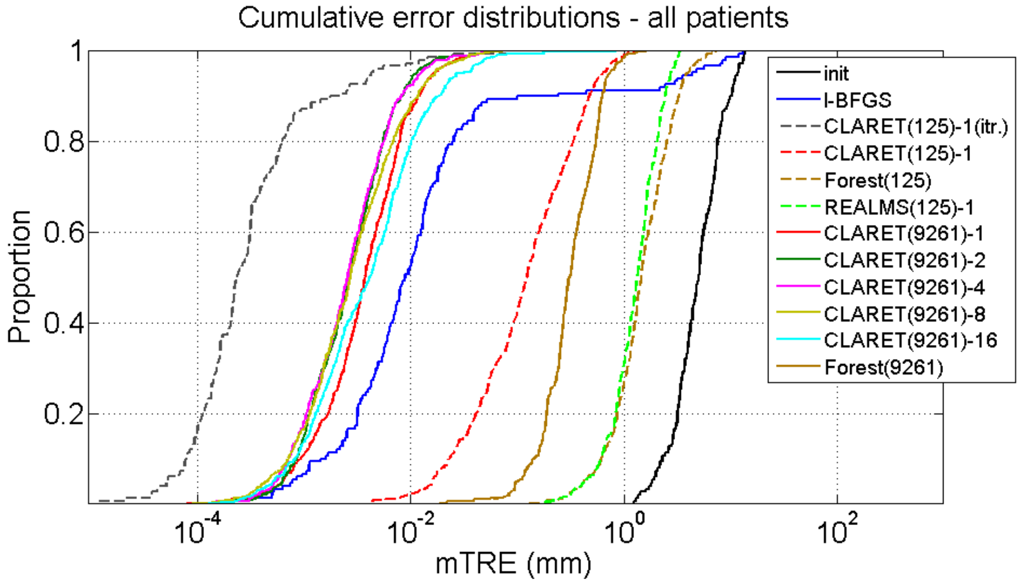

Figure 8.1.1: Registration accuracy distribution comparisons between the proposed and the optimization methods. The numbers in the parentheses are the numbers of training images used. The numbers that follow the dashed line are the numbers of partitions used for localized regression learning.

8.1.1 Iterative vs. Noniterative CLARET

As shown in the Section 5.5.2 and in the figure below (Fig. 8.1.1), the iterative CLARET with 125 training images, CLARET(125)-1(itr), greatly reduced the error from the noniterative global CLARET, CLARET(125)-1. The iterative version of CLARET showed the best registration accuracy among all of the proposed methods including the traditional optimization-based method (l-BFGS). However, due to its iterative nature, its computation time is far greater than those noniterative ones. See Table 8.1.

(ms)	CLARET(125)-1(itr.)	CLARET(9261)-1	CLARET(9261)-4	REALMS(125)-1	l-BFGS
Time	1950 ± 740	1.22 ± 0.62	80.65 ± 4.61	11.39 ± 0.73	23760 ± 14410

Table 8.1: Registration time between the proposed and the optimization methods.

When using many training images, e.g., 9261 training images, forest classification showed an improvement by a factor of four, but the noniterative CLARET, CLARET(9261), showed an improvement by a factor of ten. This indicates that CLARET's registration accuracy can be greatly improved not only by iterating the method but also by using a larger training set. This property supports the noniterative CLARET to do efficient and accurate registration at treatment time. As the computation time stated in Table 8.1, noniterative CLARETs are at least 20-fold faster than the iterative one.

8.1.2 Global vs. Local CLARET (noniterative)

Registration results of the noniterative CLARETs with various numbers of partitions have also been compared. As shown in Fig. 8.1.1, when using 9261 training images CLARET with 4 partitions, CLARET(9261)-4, yields more accurate registrations than the global CLARET, CLARET(9261)-1. Although the local CLARET is slower than the global CLARET due to the additional forest classification, the whole computation is still real-time (see Table 8.1).

8.1.3 REALMS vs. Others

As for REALMS registration using 125 training images, REALMS(125)-1, its accuracy is between the forest classification and the noniterative CLARET, CLARET(125)-1. The sources of the REALMS's insufficient accuracy may come from the following two points: (1) The non-convexity of the leave-one-out residual (Eq. 6.2.1 and Eq. 6.2.2) prevents the optimization from finding the optimal metric and kernel width. (2) Overfitting. REALMS's learned metric may overfit to the training set since it has too many parameters (weightings on the x-ray intensities) embedded in a nonlinear regression model. However, as will be stated in the next section, REALMS registration on the real tests actually works pretty well compared to results on the synthetic tests. Further discussion on this phenomenon will be provided in the next section.

8.2 Real Tests

Registration accuracy between the proposed methods and the optimization-based method have also been compared using the 6 real datasets. Largely, the results are consistent with the synthetic tests. In particular, in this real test, one new hybrid method was compared. The hybrid method used the optimization-based method but initialized from the results yielded by the global noniterative CLARET (CLARET+l-BFGS). Results of this hybrid method barely reduced errors from the optimization-based method that initialized from the mean shape (l-BFGS). This suggests that the error metrics used for the optimization-based method, namely Sum of Squared Difference (SSD), may not be valid. Future work should investigate the potential of error metrics for this 2D/3D registration problem.

Another noticeable mismatch between the synthetic tests and the real tests is in the REALMS results. REALMS performed worse in the synthetic tests but had shown pretty good accuracy on the real datasets. One potential reason of this phenomenon is that in the real datasets, there might be a greater nonlinear component between the projection intensity and the deformation parameters. As a result, REALMS, which uses the nearest neighbor approximation, can capture this higher nonlinear variation better.

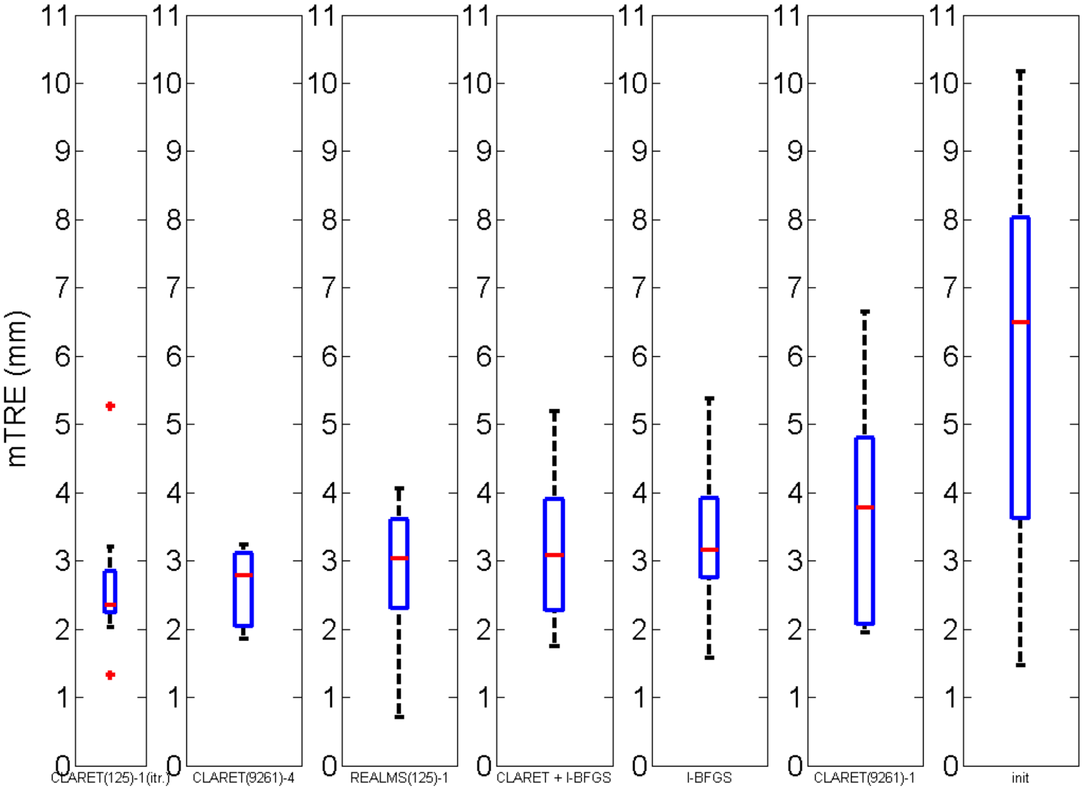

Figure 8.2.1: Box plots of the registration accuracy on real datasets for the proposed and the optimization-based method.

8.3 Which One is the Winning Method?

Which one is the winning method? It's a hard call and should depend on your target applications. Table 8.2 lists rankings in terms of accuracy and computation time for the five methods - CLARET(125)-1(itr.), CLARET(9261)-1, CLARET(9261)-4, REALMS(125)-1, and l-BFGS based on the synthetic tests. Generally speaking, if your target application does not require a real-time (<100 ms/registration) speed, the iterative CLARET may be your best choice since it yields the best accuracy. If your target application does require a real-time speed, you may consider using the local CLARET trained from a large training set, CLARET(9261)-4, as it can be computed in real time and still yield the second best accuracy. If your target application requires a higher standard of computation speed (e.g., <30 ms/registration), like fluoroscopy-to-CT registrations in the interventional surgery (Markelj *et al.* (2012); Roth *et al.* (2004)), you may consider using the global CLARET trained from a large training set, CLARET(9261)-1, as your registration method. If your target application requires a higher standard of computation speed (e.g., <30 ms/registration) and the image feature to transformation relation is known to be nonlinear, you may consider using REALMS to do registration since it uses a real-time and nonlinear registration model. Finally, if there is no a priori information of the object's transformation, you should use one of the traditional optimization-based methods (parametric PCA-based (Li *et al.* (2011a)), FFD (Rueckert *et al.* (2006, 1999)); non-parametric diffusion-based (Vercauteren *et al.* (2009); Horn and Schunck (1981)), elastic-based (Rohr *et al.* (2001)), and fluid-based (Christensen *et al.* (1996a); Beg *et al.* (2005)) registrations) since they do not require a training set.

	CLARET(125)-1(itr.)	CLARET(9261)-1	CLARET(9261)-4	REALMS(125)-1	l-BFGS
Accuracy	+++++	+++	++++	+	++
Time	++	+++++	+++	++++	+

Table 8.2: Which one is the winning method? Comparisons by rankings: the one has a higher ranking will have more '+' signs.

In application to abdominal IGRT, Zhao *et al.* (2013) recently developed a localized REALMS method that applies global-to-local metrics for parameter refinement. The method has shown an improved registration accuracy over the global REALMS with real-time computation.

Chapter 9: Conclusions and Discussion

Regression learning provides a novel solution to image registration problems, especially for registration between different image dimensions where the information of the underlying transformation is missing in the image with the lower dimension. By learning regressions between the transformations (by few global parameters) and the covarying image features, it can better estimate the latent transformations from the target image features. Per results shown in Ch. 8, regression learning for 2D/3D image registration in IGRT has been shown to be more robust, more accurate, and more efficient than the traditional optimization-based approach.

Revisit Thesis and Contributions

The contributions of this dissertation are the following:

(1) The development of four regression-learning-based 2D/3D image registration methods for image-guided radiation therapy.

a. CLARET (Correction via Limited-Angle Residues in External Beam Therapy)

At training time, the method learns linear regressions between the deformation parameters and the covarying x-ray intensity residues. The linear regression is learned to fit the simulated training set generated from the patient's treatment-planning CT images. At treatment time, it iteratively applies the learned regressions on previously unseen x-ray images to estimate the deformations.

Clinical usage of the iterative CLARET: if your target application does not require >30 registrations per second, the iterative CLARET may be your best choice since it yields the best accuracy. For 2D/3D registrations in lung IGRT, it yields an average error of 2.3 mm in 3D.

b. REALMS (Registration Efficiency and Accuracy through Learning Metric on Shape)

At training time, the method learns nonlinear kernel regressions between the deformation parameters and the covarying x-ray intensities. The kernels are Gaussian kernels that measure deformation similarities from projection distances with given kernel widths. Projection distances are equipped with Riemannian metrics. The method learns the projection distance metrics and the kernel widths that best fit to the simulated training set generated from the patient's treatment-planning CT images (by a linear regression approximation or an LOO training). At treatment time, it applies the learned regressions on previously unseen x-ray images to estimate the deformations without iteration.

Clinical usage: If your target application requires >30 registrations per second and the image feature to transformation relation is known to be nonlinear, you may consider using REALMS. For 2D/3D registrations in lung IGRT, it yields an average error of 3 mm in 3D.

c. L-CLARET (Local CLARET)

At training time, the method learns multiple locally-linear regressions between the deformation parameters and the covarying x-ray intensities. Each linear regression is learned to fit a local deformation neighborhood in the simulated training set generated from the patient's treatment-planning CT images. The deformation neighborhoods are recursively separated by normalized graph cuts with a given number of partitions. At treatment time, the method classifies a previously unseen x-ray image to its closest deformation neighborhood by a trained decision forest based on the optimized visual features and their thresholds. Then it uses the learned regressions in the chosen neighborhood on the x-ray intensities to estimate the deformations without iteration.

Clinical usage of the noniterative CLARET: If your target application require >10 registrations per second, you may consider using the local CLARET trained from a large training set as your registration method. If your target application requires >30 registrations per second, you may consider using the global CLARET trained from a large training set as your registration method. For 2D/3D registrations in lung IGRT, it yields an average error of 2.7 mm in 3D.

(2) The implementation of CLARET, L-CLARET, and REALMS into software. Code is available under *http://www.cs.unc.edu/∼cchou*

(3) The iterative versions of the four methods in **(1)**: enhancing registration accuracy by iterative estimation.

(4) The development of scattering removal and intensity correction on the digitally reconstructed radiographs (DRR) and the treatment-time radiographs to allow commensurate intensity comparison.

(5) The evaluation of CLARET, L-CLARET, REALMS and L-REALMS for lung IGRT with simulated and real patient cone-beam radiographs, including comparisons to an optimization-based 2D/3D registration approach.

(6) The evaluation of CLARET for head-and-neck IGRT with simulated and real patient NST radiographs.

Regression learning methods stated in contributions **(1)-(4)** have been studied in this dissertation. Linear and nonlinear regression learning between the transformation and the image features have been detailed in Ch. 5 and Ch. 6. Regression learning from the global and local training sets have been detailed in Ch. 7. Iterative and non-iterative regression estimation has also been detailed among those chapters. Evaluations of the proposed methods stated in contributions **(5)-(6)** have also been studied individually in Ch. 5-7 and comparatively in Ch. 8.

On the basis of these contributions, I have established the following thesis:

Thesis: *Regression learning provides a new solution to the image registration problem. Learning patient-specific intensity-to-shape regressions allows efficient, accurate, and robust 2D/3D image registration for image-guided radiation therapy.*

Problems and Future Work

How should you select the appropriate regression learning method for your target problem? The choice among the above methods should be decided case by case. One can get some clues by answering the following questions: are the target transformation parameters linear or nonlinear to the image features? Linear for all the training data, or linear for some local training set? If the transformation-to-image-feature relationship is linear, does your application particularly require computation speed (faster non-iterative approach) or does it particularly require accuracy (iterative approach for higher accuracy)?

Generally speaking, for medical images, image features are massive. As a result, with a simple multivariate linear model it is easy to find a solution in this high dimensional feature space that fits well to the training set. Using a more regularized optimization like lasso (Tibshirani (1994)), the linear model can yield a meaningful solution. However, if the image feature space of choice is low-dimensional (e.g., landmark positions), then a linear model may not have a solution. In that case, one could use a more complex nonlinear model to understand the relationship (Ch. 6), or one could use locally-linear model to understand the relationship efficiently (Ch. 7). The former takes $O(N^2)$ time for training and $O(N)$ time for registration), and the latter takes $O(N)$ time for both training and registration. N is the number of training images.

However, regression learning does not work for all image registration problems, even for the patient-specific problem that has been the target of the research of this dissertation. The limitation of this regression learning approach comes from five issues:

1. *Lack of training data*: if there are no prior transformations from which to learn, regression learning is not applicable.

2. *Target outside of the training space*: if the target transformation is not in the prior training transformation space, the method will produce inaccurate estimation.

3. *Inadequate image features to learn*: if the image features extracted from the training and target images are "variant" even under the same transformation, the proposed regression learning would not work. For example, in high-noise imaging situations like fluoroscopy in intervention surgery, it is key to select image features that are invariant between training and target times.

4. *Modeling error*: if the transformation-to-image-feature relationship is linear, fitting a nonlinear model will have an overfitting problem. On the other hand, if the transformation-to-image-feature relationship is nonlinear, a linear model will underfit the data.

5. *Projection resolution mismatch*: Limited by the CT resolution, which usually has a resolution no better than 1 mm^3, the DRRs simulated from the CT images can only have a resolution of 1 mm^2. However, the real projection images usually

have a better resolution of 0.3 mm^2. This mismatch in resolution will introduce errors in the regression inference.

Future work beyond this dissertation should try to overcome the aforementioned limitations. For examples, for (1), one can transport other patient's training data and build a training space for that patient or use nearest-neighbor search to select "nearest" patients and use the patients' training spaces as the patient's training space. For (2), one can online train the space close to the target image to extend the training space on the fly. For (3), one can implement an feature selection process that selects representable image features that are invariant in both training and target images under the same transformation. For (4), one can heuristically build a model and increase or decrease its complexity to obtain the optimal model. For (5), one can downsample and correspondingly blur the resolution for the real projection images to yield more commensurate intensity comparisons for regression inference.

BIBLIOGRAPHY

Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., and Rueckert, D. (2009). Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, **46**(3), 726–738.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, **38**(1), 95–113.

Bajcsy, R. and Broit, C. (1982). Matching of deformed images. In *Sixth International Conference on Pattern Recognition*, pages 351–353.

Bajcsy, R. and Kovacic, S. (1989). Multiresolution elastic matching. *Computer vision, graphics, and image processing*, **46**(1), 1–21.

Beg, M. F., Miller, M. I., Trouve, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, **61**, 139–157.

Bellman, R. (1986). Dynamic programming and lagrange multipliers. *The Bellman Continuum: A Collection of the Works of Richard E. Bellman*, page 49.

Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **24**(4), 509–522.

Bhatia, K. K., Hajnal, J. V., Puri, B. K., Edwards, A. D., and Rueckert, D. (2004). Consistent groupwise non-rigid registration for atlas construction. In *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pages 908–911 Vol. 1.

Blais, G. and Levine, M. (1995). Registering multiview range data to create 3d computer objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **17**(8), 820–824.

Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **11**(6), 567–585.

Breiman, L. (2001). Random forests. *Mach. Learn.*, **45**(1), 5–32.

Bro-Nielsen, M. and Cotin, S. (1996). Real-time volumetric deformable models for surgery simulation using finite elements and condensation. In *Computer graphics forum*, volume 15, pages 57–66. Wiley Online Library.

Broit, C. (1981). Optimal registration of deformed images.

Burnet, N. G., Thomas, S. J., Burton, K. E., and Jefferies, S. J. (2004). Defining the tumour and target volumes for radiotherapy. *Cancer Imaging*, **4(2)**, 153–161.

Capital Radiation Therapy (2013). http://www.capitalmedical.com/services-technology/medical-services-technology/capital-radiation-therapy/.

Chen, Q.-s., Defrise, M., and Deconinck, F. (1994). Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **16**(12), 1156–1168.

Chou, C., Frederick, C., Chang, S., and Pizer, S. (2010a). A Learning-Based patient repositioning method from Limited-Angle projections. In J. A. et al., editor, *Brain, Body and Machine*, volume 83 of *Advances in Soft Computing*, pages 83–94. Springer Berlin/Heidelberg.

Chou, C.-R. and Pizer, S. (2013). Local regression learning via forest classification for 2d/3d deformable registration. In *MICCAI workshop on Medical Computer Vision (MCV)*. to appear.

Chou, C.-R. and Pizer, S. M. (2012). Real-time 2d/3d deformable registration using metric learning. In *MICCAI workshop on Medical Computer Vision (MCV)*, pages 1–10.

Chou, C.-R., Pizer, S. M., Frederick, C. B., and Chang, S. X. (2010b). Igrt via machine learning from limited angle projection images. *Medical Physics*, **37**, 3332.

Chou, C.-R., Frederick, B., Liu, X., Mageras, G., Chang, S., and Pizer, S. (2011a). Claret: A fast deformable registration method applied to lung radiation therapy. In *Fourth International (MICCAI) Workshop on Pulmonary Image Analysis*, pages 113–124.

Chou, C.-R., Chang, S. X., Frederick, C. B., and Pizer, S. M. (2011b). Limited projection image based deformable three-dimensional registration for lung and head and neck. In *Cancer Imaging and Radiation Therapy Symposium*. To appear.

Chou, C.-R., Frederick, B., Chang, S., and Pizer, S. (2011c). Real-time deformable registration using multi-pixel x-ray technology for image-guided lung radiotherapy. *Medical Physics*, **38**, 3818.

Chou, C.-R., Frederick, C. B., Chang, S., and Pizer, S. (2012). Mo-f-bra-05: Real-time 3d tumor localization for lung igrt using a single x-ray projection. *Medical Physics*, **39**, 3875.

Chou, C.-R., Frederick, B., Mageras, G., Chang, S., and Pizer, S. (2013). 2d/3d image registration using regression learning. *Computer Vision and Image Understanding*, **117**(9), 1095 – 1106.

Christensen, G., Rabbitt, R., and Miller, M. (1996a). Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, **5**(10), 1435–1447.

Christensen, G. E. (1994). *Deformable shape models for anatomy*. Ph.D. thesis, Washington University Saint Louis, Mississippi.

Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1994a). 3d brain mapping using a deformable neuroanatomy. *Physics in medicine and biology*, **39**(3), 609.

Christensen, G. E., Miller, M. I., and Vannier, M. (1994b). A 3d deformable magnetic resonance textbook based on elasticity. In *Proceedings of the American association for artificial intelligence, symposium: applications of computer vision in medical image processing*. Citeseer.

Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1996b). Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, **5**, 1435–1447.

Chui, H. and Rangarajan, A. (2003). A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, **89**, 114 – 141.

Collins, D. L. and Evans, A. C. (1997). Animal: validation and applications of nonlinear registration-based segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, **11**(08), 1271–1294.

Cootes, T., Marsland, S., Twining, C., Smith, K., and Taylor, C. (2004). Groupwise diffeomorphic non-rigid registration for automatic model building. In T. Pajdla and J. Matas, editors, *Computer Vision - ECCV 2004*, volume 3024 of *Lecture Notes in Computer Science*, pages 316–327. Springer Berlin Heidelberg.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(6), 681–685.

Criminisi, A., Shotton, J., and Konukoglu, E. (2011). Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical Report MSR-TR-2011-114, Microsoft Research.

D'Agostino, E., Maes, F., Vandermeulen, D., and Suetens, P. (2003). A viscous fluid model for multimodal non-rigid image registration using mutual information. *Medical Image Analysis*, **7**(4), 565 – 575.

Davatzikos, C., Vaillant, M., Resnick, Susan M.and Prince, J. L. L. S., and Bryan, R. N. (1996). A computerized approach for morphological analysis of the corpus callosum. *Journal of Computer Assisted Tomography*, **20**(1), 88–97.

De Boor, C., De Boor, C., De Boor, C., and De Boor, C. (1978). *A practical guide to splines*, volume 27. Springer-Verlag New York.

Dupuis, P., Grenander, U., and Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, **56**(3), 587.

Fischer, B. and Modersitzki, J. (2002). Fast diffusion registration. *Contemporary Mathematics*, **313**, 117–128.

Fornefett, M., Rohr, K., and Stiehl, H. S. (2001). Radial basis functions with compact support for elastic registration of medical images. *Image and Vision Computing*, **19**(1), 87–96.

Fox, N. C., Crum, W. R., Scahill, R. I., Stevens, J. M., Janssen, J. C., and Rossor, M. N. (2001). Imaging of onset and progression of alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *The Lancet*, **358**(9277), 201 – 205.

Frederick, B., Lalush, D., , and Chang, S. (2010). Th-c-201c-10: Registration using nanotube stationary tomosynthesis: Comparison of 3d/3d to 3d/2d methods. *Medical Physics*, **37**, 3460.

Fu, D. and Kuduvalli, G. (2008). A fast, accurate, and automatic 2d–3d image registration for image-guided cranial radiosurgery. *Medical physics*, **35**, 2180.

Gaens, T., Maes, F., Vandermeulen, D., and Suetens, P. (1998). Non-rigid multimodal image registration using mutual information. In W. Wells, A. Colchester, and S. Delp, editors, *Medical Image Computing and Computer-Assisted Interventation*, volume 1496 of *Lecture Notes in Computer Science*, pages 1099–1106. Springer Berlin Heidelberg.

Gee, J. C., Haynor, D. R., Le Briquer, L., and Bajcsy, R. K. (1997). Advances in elastic matching theory and its implementation. In *CVRMed-MRCAS'97*, pages 63–72. Springer.

Hajnal, J. V. and Hill, D. L. (2010). *Medical image registration*. CRC press.

Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *ARTIFICAL INTELLIGENCE*, **17**, 185–203.

Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M. A., and van Ginneken, B. (2009). Multi-atlas-based segmentation with local decision fusion - application to cardiac and aortic segmentation in ct scans. *Medical Imaging, IEEE Transactions on*, **28**(7), 1000–1010.

Jenkinson, M. and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, **5**(2), 143–156.

Knaan, D. and Joskowicz, L. (2003). Effective intensity-based 2d/3d rigid registration between fluoroscopic x-ray and ct. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*, pages 351–358. Springer.

Kruskal, J. (1964). Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, **29**, 1–27.

Lewis, J. (1995). Fast normalized cross-correlation. In *Vision interface*, volume 10, pages 120–123.

Li, R., Jia, X., Lewis, J. H., Gu, X., Folkerts, M., Men, C., and Jiang, S. B. (2010). Real-time volumetric image reconstruction and 3d tumor localization based on a single x-ray projection image for lung cancer radiotherapy. *Medical Physics*, **37**, 2822–2826.

Li, R., Lewis, J. H., Jia, X., Gu, X., Folkerts, M., Men, C., Song, W. Y., and Jiang, S. B. (2011a). 3d tumor localization through real-time volumetric x-ray imaging for lung cancer radiotherapy. *Medical Physics*, **38**, 2783–2794.

Li, R., Lewis, J. H., Jia, X., Zhao, T., Liu, W., Wuenschel, S., Lamb, J., Yang, D., Low, D. A., and Jiang, S. B. (2011b). On a pca-based lung motion model. *Physics in medicine and biology*, **56**(18), 6009.

Liu, X., Davis, B., Niethammer, M., Pizer, S., and Mageras., G. (2010). Prediction-driven respiratory motion atlas formation for 4d image- guided radiation therapy in lung. In *MICCAI'10 Pulmonary Image Analysis Workshop*.

Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., and Joshi, S. (2006). Multi-modal image set registration and atlas formation. *Medical Image Analysis*, **10**(3), 440–451.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on*, **16**(2), 187–198.

Maltz, J. S., Sprenger, F., Fuerst, J., Paidi, A., Fadler, F., and Bani-Hashemi, A. R. (2009). Fixed gantry tomosynthesis system for radiation therapy image guidance based on a multiple source x-ray tube with carbon nanotube cathodes. *Medical Physics*, **36**, 1624–1636.

Markelj, P., Tomazevic, D., Likar, B., and Pernus, F. (2012). A review of 3d/2d registration methods for image-guided interventions. *Medical Image Analysis*, **16**(3), 642–661.

Miller, M. I. and Younes, L. (2001). Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, **41**(1-2), 61–84.

Miller, M. I., Trouve, A., and Younes, L. (2002). On the metrics and euler-lagrange equations of computational anatomy. *Annual review of biomedical engineering*, **4**(1), 375–405.

Modersitzki, J. (2004). *Numerical Methods for Image Registration*. Oxford University Press.

Niethammer, M., Huang, Y., and Vialard, F.-X. (2011). Geodesic regression for image time-series. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 655–662. Springer.

Nocedal, J. (1980). Updating quasi-newton matrices with limited storage. *Mathematics of Computation*, **35**, 773–782.

Peters, G. and Wilkinson, J. H. (1970). The least squares problem and pseudo-inverses. *The Computer Journal*, **13**(3), 309–316.

Pickering, M. R., Muhit, A. A., Scarvell, J. M., and Smith, P. N. (2009). A new multi-modal similarity measure for fast gradient-based 2d-3d image registration. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 5821–5824. IEEE.

Pluim, J. P. W., Maintz, J., and Viergever, M. (2003). Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, **22**(8), 986–1004.

Reddy, B. S. and Chatterji, B. N. (1996). An fft-based technique for translation, rotation, and scale-invariant image registration. *Image Processing, IEEE Transactions on*, **5**(8), 1266–1271.

Roche, A., Malandain, G., Pennec, X., and Ayache, N. (1998). The correlation ratio as a new similarity measure for multimodal image registration. In W. Wells, A. Colchester, and S. Delp, editors, *Medical Image Computing and Computer-Assisted Interventation*, volume 1496 of *Lecture Notes in Computer Science*, pages 1115–1124. Springer Berlin Heidelberg.

Rohlfing, T., Russakoff, D. B., Murphy, M. J., and Maurer Jr, C. R. (2002). Intensity-based registration algorithm for probabilistic images and its application for 2d to 3d image registration. In *Medical Imaging 2002*, pages 581–591. International Society for Optics and Photonics.

Rohr, K., Stiehl, H., Sprengel, R., Buzug, T., Weese, J., and Kuhn, M. (2001). Landmark-based elastic registration using approximating thin-plate splines. *Medical Imaging, IEEE Transactions on*, **20**(6), 526–534.

Roth, M., Dotter, M., Burgkart, R., and Schweikard, A. (2004). Fast intensity-based fluoroscopy-to-ct registration using pattern search optimization. In *International Congress Series*, volume 1268, pages 165–170. Elsevier.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on*, **18**(8), 712–721.

Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V., and Hammers, A. (2006). Diffeomorphic registration using b-splines. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 702–709. Springer.

Russakoff, D., Rohlfing, T., and Maurer, C. (2003). Fast intensity-based 2D-3D image registration of clinical data using light fields. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, volume 1, pages 416–422.

Russakoff, D. B., Rohlfing, T., Mori, K., Rueckert, D., Ho, A., Adler, J. R., and Maurer, C. R. (2005). Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2d-3d image registration. *IEEE Transactions on Medical Imaging*, **24**, 1441–1454.

Sarrut, D. and Clippe, S. (2001). Geometrical transformation approximation for 2d/3d intensity-based registration of portal images and ct scan. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*, pages 532–540. Springer.

Sherouse, G. W., Novins, K., and Chaney, E. L. (1990). Computation of digitally reconstructed radiographs for use in radiotherapy treatment design. *International Journal of Radiation Oncology* Biology* Physics*, **18**(3), 651–658.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**, 888–905.

Thirion, J.-P. (1998). Image matching as a diffusion process: an analogy with maxwell's demons. *Medical image analysis*, **2**(3), 243–260.

Tibshirani, R. (1994). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.

Trouve, A. (1995). An infinite dimensional group approach for physics based models in pattern recognition. *Preprint*.

Vandemeulebroucke, J., Kybic, J., Sarrut, D., and Clarysse, P. (2009). Breathing thorax model from ct images for use in radiotherapy.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, **45**(1, Supplement 1), S61 – S72.

Wang, L., Beg, F., Ratnanather, T., Ceritoglu, C., Younes, L., Morris, J. C., Csernansky, J. G., and Miller, M. I. (2007). Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the alzheimer type. *Medical Imaging, IEEE Transactions on*, **26**(4), 462–470.

Weese, J., Buzug, T. M., Lorenz, C., and Fassnacht, C. (1997). An approach to 2d/3d registration of a vertebra in 2d x-ray fluoroscopies with 3d ct images. In *CVRMed-MRCAS'97*, pages 119–128. Springer.

Weinberger, K. and Tesauro, G. (2007). Metric learning for kernel regression. In *Eleventh international conference on artificial intelligence and statistics*, pages 608–615.

Wu, M., Rosano, C., Lopez-Garcia, P., Carter, C. S., and Aizenstein, H. J. (2007). Optimum template selection for atlas-based segmentation. *NeuroImage*, **34**(4), 1612–1618.

Yao, J. and Taylor, R. (2003). Assessing accuracy factors in deformable 2d/3d medical image registration using a statistical pelvis model. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1329–1334. IEEE.

Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C., and Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, **31**(3), 1116–1128.

Zeidan, O. A., Langen, K. M., Meeks, S. L., Manon, R. R., Wagner, T. H., Willoughby, T. R., Jenkins, D. W., and Kupelian, P. A. (2007). Evaluation of image-guidance protocols in the treatment of head and neck cancers. *International Journal of Radiation Oncology* Biology* Physics*, **67**(3), 670–677.

Zhao, Q., Chou, C.-R., Mageras, G., and Pizer, S. (2013). Local metric learning in 2d/3d deformable registration with application in the abdomen. Manuscript submitted for a journal review.

Zhong, J. and Qiu, A. (2010). Multi-manifold diffeomorphic metric mapping for aligning cortical hemispheric surfaces. *Neuroimage*, **49**(1), 355–365.

Zollei, L., Grimson, E., Norbash, A., and Wells, W. (2001). 2d-3d rigid registration of x-ray fluoroscopy and ct images using mutual information and sparsely sampled histogram estimators. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–696. IEEE.