Robust and Fast 2D/3D Image Registration using Regression Learning

Chen-Rui Chou [a], Brandon Frederick [b], Gig Mageras [d], Sha Chang [b,c],
Stephen Pizer [a,b,c]

[a] *Departments of Computer Science,* [b] *Biomedical Engineering,* [c] *Radiation Oncology,*

University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[d] *Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA*

Abstract

In computer vision and image analysis, image registration between 2D projections and a 3D image while obtaining high accuracy and real-time computation is challenging. In this paper, we propose a novel method that can speedily detect the object's 3D rigid motion or deformation from a small set of its 2D projection images. The method consists of two stages: registration and pre-registration learning. In the registration stage, it iteratively estimates the motion/deformation parameters based on the current intensity residue between the target projection and the projection of the estimated 3D image using learned linear operators. The linear operators are learned in the pre-registration learning stage: First, it builds a low-order parametric model of the image region's motion/deformation shape space from its prior 3D images. Second, using learning-time samples produced from the 3D images, it formulates the relationships between the model parameters and the co-varying 2D projection intensity residues by multi-scale linear regressions. The calculated multi-scale regression matrices give the coarse to fine linear operators used in estimating the model parameters from the 2D projection intensity residues in the registration. The method's application to Image-guided Radiation Therapy (*IGRT*), called *CLARET* (Correction via Limited-Angle Residues in External Beam Therapy), requires only a few seconds and has given good results in localizing a tumor under rigid motion in the head and neck and under respiratory deformation in the lung using a small set of treatment-time imaging 2D projections.

Key words: 2D/3D Registration, Regression, Machine Learning, IGRT, Radiation Therapy

1. Introduction

In a variety of situations in therapy guidance in medicine, there is a planning-time 3D image taken and there is a set of treatment-time 2D images taken that are used to find the 2D/3D geometric transformation relating the treatment-time layout to the planning-time layout. Current available 2D/3D registration methods, Russakoff et al. [1, 2], Khamene et al. [3], Munbodh et al. [4], Li et al. [5, 6], find that geometric transformation that optimizes an objective function consisting of a match term to the 2D image data and a regularization term. As a fast optimization often requires many evaluations of the function's Jacobian, the optimization-based registration methods are structurally slow. Also, the mismatch in the registration dimensionality often introduces a non-convex objective function which is prone to make the optimization fall into local minima. We have sought a method that is both fast and not based on optimization at registration time.

In a way similar to the face alignment algorithm *AAM* (Active Appearance Model) in Cootes et al. [7] and the efficient tracking scheme in Jurie and Dhome [8], we seek a linear operator $\mathbf{M}$, calculated by linear regression, that when iteratively applied to intensity differences (*residue*) $\mathbf{R}$ between projections of presently estimated 3D image $\mathbf{I}_{3D}^t$ and the measured images Ψ, yields parameters $\mathbf{C}$ of the transformation T needed to lessen the residue.

$$\mathbf{C}^t = \mathbf{M} \cdot \mathbf{R}_{2D}^t \qquad (1.1)$$

$$\mathbf{I}_{3D}^{t+1} = \mathbf{I}_{3D}^t \circ T(\mathbf{C}^t) \qquad (1.2)$$

The registration process in eqs. 1.1 and 1.2 requires no optimizations, and therefore it can support efficient registration. Different from the AAM, our linear operator $\mathbf{M}$ estimates 3D transformation parameters from 2D projection intensity residues $\mathbf{R}$ for the 2D/3D registration.

We describe our method in detail in the following sequence: First, we describe our 2D/3D registration framework and our efficient approximation of the shape parameter $\mathbf{C}$ in section 2. In section 3, we describe how we obtain low-order parametrization for the rigid motion and deformation shape space. In section 4, we describe our regression learning to calculate the linear operator $\mathbf{M}$ and an efficient multi-scale learning scheme. In section 5, we describe how we generate commensurate projection intensities for our linear operator to work. In section 6, we describe our medical application. In sections 7 and 8, we show and then discuss our rigid and non-rigid registration results.

2. 2D/3D Registration

In this section, we first describe the general framework of our 2D/3D image registration method. Second, we describe our approach to do efficient registration within this framework.

2.1. General 2D/3D Registration

The goal of the 2D/3D registration is to match a transformed 3D source image I to a set of target 2D projections Ψ. We denote the projection intensity at pixel location $\mathbf{x} = (x_1, x_2)$ from a projection angle θ as $\Psi(\mathbf{x}; \theta)$. The registration can be formulated as an iterative process. Let $I(t)$ denote the gray-scale 3D image at iteration t. At iteration t, the 3D image region's motion/deformation parameters $\mathbf{C}(t)$ define a geometric transformation $T(\mathbf{C}(t))$ in a shape space determined from the 3D images. We obtain $\mathbf{C}(t)$ from projection intensity residues $\mathbf{R}$ between the target 2D projections $\Psi(\mathbf{x}; \theta)$ and the computed projections $\mathbf{P}(\mathbf{x}, I(t); \theta)$ of the transformed 3D source image $I(t) = I(t-1) \circ T(\mathbf{C}(t-1))$. We denote the computed projection intensity at pixel location $\mathbf{x} = (x_1, x_2)$ from a projection angle θ from a 3D image $I(t)$ in iteration t as $\mathbf{P}(\mathbf{x}, I(t); \theta)$.

$$\mathbf{R}[\Psi(\mathbf{x}; \theta), \mathbf{P}(\mathbf{x}, I(t); \theta)] = \Psi(\mathbf{x}; \theta) - \mathbf{P}(\mathbf{x}, I(t); \theta) \qquad (2.1)$$

The projection operator $\mathbf{P}$ is formulated by a simulation of the imaging process. For example, in the medical literature, to simulate a 3D image's x-ray projections from its 3D volume, we do ray casting to simulate the photon attenuation for a given imaging geometry (figure 2.1).
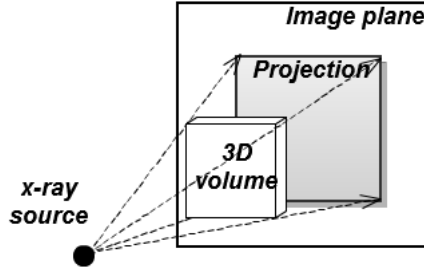

Figure 2.1: An x-ray projection is simulated by ray casting on a 3D image volume. The dashed lines and the arrows indicate the ray directions.

One way to calculate $\mathbf{C}$ is to optimize it with respect to certain measure ρ of the intensity residue $\mathbf{R}$.

$$\mathbf{C}(t) = \underset{\mathbf{C}}{argmin} \left\| \mathbf{R}[\Psi(\mathbf{x}), \mathbf{P}(\mathbf{x}, I(t) \circ T(\mathbf{C}))] \right\|_{\rho} \qquad (2.2)$$

Computations iteratively accomplishing this optimization are slow in general. Moreover, the iterations may easily converge to local minima since the energy functional in eq. 2.2 is not convex.

2.2. Efficient Linear Approximation of $\mathbf{C}$

We propose a different method to calculate $\mathbf{C}$ using multi-scale linear operators $\mathbf{M}$. At iteration t of registration, our method estimates the motion/deformation

3

parameters $\mathbf{C}(t)$ by applying a linear operator $\mathbf{M}_s$ of scale s to the current intensity residues $\mathbf{R}[\Psi(\mathbf{x}), \mathbf{P}(\mathbf{x}, I(t))]$, defined as the concatenation over all of the projection angles θ of the residues $\mathbf{R}[\Psi(\mathbf{x}; \theta), \mathbf{P}(\mathbf{x}, I(t); \theta)]$. That is,

$$\mathbf{C}(t) = \mathbf{R}[\Psi(\mathbf{x}), \mathbf{P}(\mathbf{x}, I(t))] \cdot \mathbf{M}_s, \text{ where } s = 1, 2, \cdots S; \, t = 1, 2, \cdots, t_{max} \quad (2.3)$$

Typically, $S = 4$ and $t_{max} \leq 10$ are satisfactory. The computation in eq. 2.3 involves only matrix multiplications by $\mathbf{M}_s$, computing the projections by $\mathbf{P}$, and subtractions. This makes the registration structurally fast. The calculation of the multi-scale linear operators $\mathbf{M}$ involves a machine learning process described in detail in section 4.

3. Shape Space Modeling

Our method limits the motion/deformation to a shape space. To allow $\mathbf{M}$ to be accurately learned, we require a low-order parametrization $\mathbf{C}$ of this shape space. We describe the shape space calculation for rigid motions and for non-rigid deformations in section 3.1 and 3.2 respectively.

3.1. Rigid Motion Modeling

Rigid motions are modeled explicitly as the variation in the Euler's six dimensional rigid space:

$$\mathbf{C} = (t_x, t_y, t_z, r_x, r_y, r_z) \quad (3.1)$$

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 described as angular rotation in degrees ($°$) about the image center successively around the world's coordinate axes x, y, z, respectively.

3.2. Deformation Modeling

We model deformations as a linear combination of a set of basis deformations calculated through PCA analysis. In our target problem, a cyclically varying set of 3D images $\{J_\tau \text{ over time } \tau\}$ are available at pre-registration learning time. From these a mean image $\bar{J} = I$ and a set of deformations ϕ_τ between J_τ and $\bar{J}$ can be computed. The basis deformations can then be chosen to be the primary eigenmodes of a PCA analysis on the ϕ_τ.

3.2.1. Deformation Shape Space and Mean Image Generation

In order to model the deformation space realistically, our method computes a deformation-based intrinsic mean as a Fréchet mean image $\bar{J}$ via an *LDDMM* (Large Deformation Diffeomorphic Metric Mapping) framework described in Beg et al. [9] on the cyclically varying set of 3D images $\{J_\tau \text{ over time } \tau\}$. The Fréchet mean $\bar{J}$, as well as the diffeomorphic deformations ϕ from the mean

$\overline{J}$ to each image J_τ, are computed using a fluid-flow distance metric d_{fluid} as described in Christensen et al. [10] and Davis et al. [11]:

$$\overline{J} \;=\; \underset{J}{argmin} \sum_{\tau=1}^{N} d_{fluid}(J, J_\tau)^2 \tag{3.2}$$

$$= \underset{J}{argmin} \sum_{\tau=1}^{N} \int_0^1 \int_\Omega ||v_{\tau,\gamma}(x)||^2 dx d\gamma + \frac{1}{\sigma^2} \int_\Omega ||\bar{J}(\phi_\tau^{-1}(x)) - J_\tau(x)||^2 dx \tag{3.3}$$

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

3.2.2. Statistical Analysis

With the diffeomorphic deformation set $\{\phi_\tau$ over $\tau\}$ calculated, our method finds a set of linear deformation basis functions ϕ_{pc}^i by PCA analysis. The scores λ_τ^i (basis function weights) for each ϕ_{pc}^i yield ϕ_τ in terms of these basis functions.

$$\phi_\tau = \overline{\phi} + \sum_{i=1}^{N} \lambda_\tau^i \cdot \phi_{pc}^i \tag{3.4}$$

We choose a subset of n eigenmodes that captures 95% of the total variation. Then we let the n basis function weights λ^i form the the n-dimensional parametrization $\mathbf{C}$.

$$\mathbf{C} = (\lambda^1, \lambda^2, \cdots, \lambda^n) \tag{3.5}$$

4. Machine Learning

With the motion/deformation's shape space we calculate linear operators $\mathbf{M}$ that correlate coarsely to finely sampled model parameters $\mathbf{C}$ with the corresponding projection intensity residue vectors $\mathbf{R}$. We describe our regression learning to calculate the linear operators $\mathbf{M}$ in section 4.1 and an efficient multiscale learning strategy in section 4.2.

4.1. Residues to Model Parameters Regression Learning

As detailed in section 4.2 we select a collection of model parameters $\{\mathbf{C}_\kappa$ over case $\kappa\}$ for learning. We use linear regression to correlate the selected modeled parameters $\mathbf{C}_\kappa$ in the κth case with the co-varying projection intensity residue

set $\{\mathbf{R}_{\kappa,\theta}$ over the projection angle $\theta\}$. $\mathbf{R}_{\kappa,\theta}(\mathbf{x})$ can be formulated as the projection intensity difference at pixel location $\mathbf{x} = (x_1, x_2)$ from a projection angle θ between the mean image (or an untransformed 3D image for the rigid case) I and the image $I \circ T(\mathbf{C}_\kappa)$ transformed with the sampled model parameter $\mathbf{C}_\kappa$:

$$\mathbf{R}_{\kappa,\theta}(\mathbf{x}) = \mathbf{P}(\mathbf{x}, I \circ T(\mathbf{C}_\kappa); \theta) - \mathbf{P}(\mathbf{x}, I; \theta) \qquad (4.1)$$

We concatenate the residues at each projection angle to formulate a residue set in a vector $\mathbf{R}_\kappa = (\mathbf{R}_{\kappa,\theta_1}, \mathbf{R}_{\kappa,\theta_2}, \cdots, \mathbf{R}_{\kappa,\theta_\Gamma})$ and build a linear regression for all cases $\kappa = 1, 2, \cdots, K$:

$$\begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_K \end{pmatrix} \approx \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \vdots \\ \mathbf{R}_K \end{pmatrix} \cdot \mathbf{M} \qquad (4.2)$$

The regression matrix $\mathbf{M}$ that gives the best estimation of the linear operators per parameter scale is computed via a pseudo-inverse:

$$\mathbf{M} = (\mathbf{R}^\mathsf{T}\mathbf{R})^{-1}\mathbf{R}^\mathsf{T}\mathbf{C} \qquad (4.3)$$

4.2. Multi-scale Learning

To enhance the regression learning ability, $\mathbf{C}$ must be sufficiently sampled to capture all the shape variations. Doing this directly needs an exponential time computation. Instead, we have designed an efficient learning scheme that learns the model parameters, from large to small scales 1 to S, to minimize the regression residual in sequence, yielding S multi-scale regression matrices $\mathbf{M}_1, \mathbf{M}_2, \cdots, \mathbf{M}_S$. At the sth scale of learning, each model parameter c in $\mathbf{C}$ is collected from the combinations of $\pm c_s$ and 0 where c_s is the extreme value of the range of parameter c at scale s. In order to have accurate estimations in the whole model domain, the selection of each c_s depends on the interpolation accuracy of $\mathbf{M}_s$ such that the union of the intervals of acceptable accuracy $\overset{S}{\underset{s=1}{\cup}} \Lambda_s^c$ for parameter c, covers the whole model domain:

$$\overset{S}{\underset{s=1}{\cup}} \Lambda_s^c \supseteq [-c_1, \ c_1] \qquad (4.4)$$

In the registration stage the calculated multi-scale linear operators are applied sequentially, from $\mathbf{M}_1$ to $\mathbf{M}_S$, to give new estimations of the model parameters from large to small scales. After evaluating the estimation accuracy for target examples of both the rigid and non-rigid types, we found that four scales ($S = 4$) with $c_s = (S - s + 1) \cdot c_1/S$ gave the required accuracy.

5. Commensurate Projection Intensity Generation

The linear operators $\mathbf{M}$ in the learning stage are not invariant to the projection intensity scales and variations caused by x-ray scatter. Therefore, our method

uses a normalization filter (section 5.1) and a following histogram matching (section 5.2) to generate commensurate intensities for the learning-time computed projections and the registration-time target projections.

5.1. Local Gaussian Normalization

To account for variations caused by x-ray scatter, we do a 2D Gaussian-weighted normalization for each pixel in the learning projections (see figure 5.1 (d)) and the target projections (see figure 5.1 (b)). To calculate the normalized value $\Psi'(\mathbf{x}; \theta)$ at a pixel location $\mathbf{x} = (x_1, x_2)$ for each projection angle θ, we 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 $\sigma'(x_1, x_2)$.

$$\mu'(x_1, x_2) = \frac{\sum_{\xi=x_1-A}^{x_1+A} \sum_{\eta=x_2-B}^{x_2+B} [G(\xi, \eta; 0, w) * \Psi(\xi, \eta)]}{(2A+1) \times (2B+1)} \tag{5.1}$$

where $*$ is the 2D convolution operator; $2A+1$ and $2B+1$, respectively, are the number of columns and rows in the averaging window centered at (x_1, x_2); the function G is a zero mean Gaussian distribution with a standard deviation w.

$$\sigma'(x_1, x_2) = \left(\frac{\sum_{\xi=x_1-A}^{x_1+A} \sum_{\eta=x_2-B}^{x_2+B} [G(\xi, \eta; 0, w) * \Psi(\xi, \eta) - \mu'(x_1, x_2)]^2}{(2A+1) \times (2B+1)} \right)^{\frac{1}{2}}$$
$$\tag{5.2}$$

$$\Psi'(x_1, x_2) = \frac{\Psi(x_1, x_2) - \mu'(x_1, x_2)}{\sigma'(x_1, x_2)} \tag{5.3}$$

We choose A, B, and w to be a few pixels to perform a local Gaussian-weighted normalization for our target problem (see section 6).

5.2. Histogram Matching

In addition, 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 on intensity achieving non-linear cumulative histogram matching within the region ω of an object of interest is applied after local Gaussian normalization. The object region ω is determined as a pixel set where their intensity values are larger than the mean value in the projection. That is, F_ω is defined by

$$F_\omega(H_f(\Psi'_{target})) \approx H_f(\Psi'_{learning}) \tag{5.4}$$

where H_f is the cumulative histogram profiling function. The histogram matched intensities $\Psi^\star_{target}$ (see figure 5.1 (c)) can be calculated through the mapping:

$$\Psi^\star_{target} = \Psi'_{target} \circ F_\omega \tag{5.5}$$
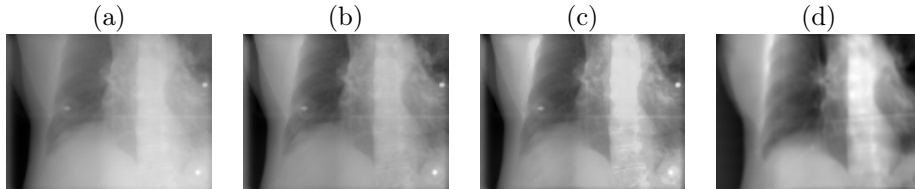
Figure 5.1: (a) A raw Cone-Beam (CB) projection (target projection), (b) a local Gaussian normalized CB projection (normalized target projection), (c) histogram matched CB projection (normalized and histogram matched target projection) and (d) a local Gaussian normalized DRR of a Fréchet mean CT (learning 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.

6. Medical Application - IGRT

Our target problem is IGRT (Image-guided Radiation Therapy). There the 3D image I is the planning CT (Computed Tomography), 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) a stationary *NST* (Nanotube Stationary Tomosynthesis) imager specified in Maltz et al. [12]. Our method's application to IGRT, called *CLARET* (Correction via Limited-Angle Residues in External Beam Therapy) has shown promise in registering the planning CT to the treatment-time imaging projections. We describe the two treatment imaging geometries in section 6.1 and CLARET's specialization for head-and-neck IGRTs and lung IGRTs in sections 6.2 and 6.3, respectively.

6.1. Treatment Imaging Geometry

6.1.1. Cone-beam CT (CBCT)

CBCT is a rotational imaging system with a single radiation source and a planar detector, and is mounted on a medical linear accelerator. This pair rotates by an angle of up to 2π during IGRT, taking projection images Ψ during traversal (figure 6.1 (a)). A limited-angle rotation provides a shortened imaging time and lowered imaging dose. For example, for a 5° rotation, it takes ~ 1 second.

6.1.2. Nanotube Stationary Tomosynthesis (NST)

NST is a stationary imaging system mounted on a medical linear accelerator that can perform imaging without interfering with treatment delivery. As illustrated in figure 6.1 (b), it consists of an arrangement of radiation sources arrayed around the treatment portal, together with a planar detector. Firing the sources in sequence produces a different 2D projection image Ψ per source. Each projection image requires ~ 200 ms.
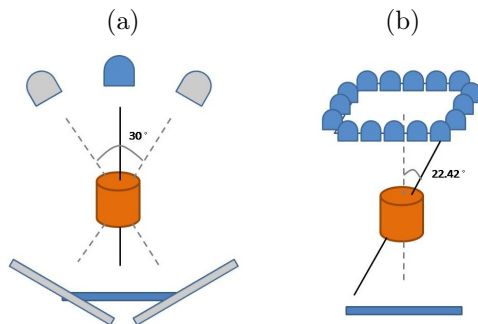
Figure 6.1: (a) The 30° CBCT geometry: rotational imaging system with tomographic angle $\theta = 15°$ (b) The NST geometry: stationary sources arrays with max tomographic angle $\theta = 22.42°$

6.2. Head-and-neck IGRT

In head-and-neck IGRT, the use of an immobilization device allows very little geometric difference between planning time and treatment time other than a rigid transformation. Therefore, in the pre-registration learning, CLARET samples clinically credible variations (± 2 cm, $\pm 5°$) in the Euler's 6-space $\mathbf{C}$ to capture the treatment-time patient's motions. With a single planning CT I of the patient the learning computed projections $\mathbf{P}(\mathbf{x}, I \circ T(\mathbf{C}); \theta)$ are generated by transformation of those credible variations $T(\mathbf{C})$ and projection from a given tomographic angle θ to the transformed 3D volume $I \circ T(\mathbf{C})$.

In the registration, CLARET iteratively applies S multi-scale linear operators, from $\mathbf{M}_1$ to $\mathbf{M}_S$ to estimate the rigid transformation parameters from the 2D intensity residues form by the difference between the target projections $\Psi^\star$ and the normalized projections computed from the presently estimated rigid transformation applied to the planning-time 3D image.

6.3. Lung IGRT

In lung IGRT, the respiratory motion introduces non-rigid transformations. In the pre-registration learning stage, a set of 10-phase RCCTs (Respiratory-correlated CTs) collected at planning time give 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 respiratory Fréchet mean image $\overline{J}$ and the deformation ϕ_τ to $\overline{J}$ corresponding to the CT J_τ are calculated via an LDDMM framework. See an example respiratory Fréchet mean image in figure 6.2 (c). The deformation basis functions ϕ_{pc} are then generated by PCA analysis on the deformation set $\{\phi_\tau$ over phase $\tau\}$. In Liu et al. [13], they have shown that a shape space with three eigenmodes adequately captures the respiratory variations experienced at treatment time. See the first two principal deformation basis functions in figure 6.3.
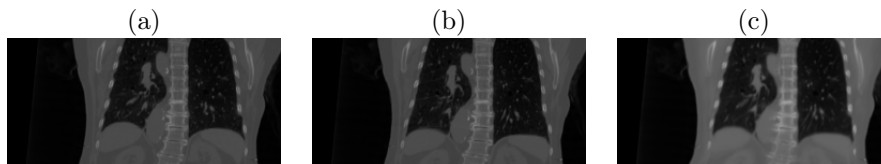
Figure 6.2: (a) Planning CT at the End-Expiration (EE) phase (b) planning CT at the End-Inspiration (EI) phase and (c) respiratory Fréchet mean CT generated via an LDDMM framework from a lung dataset .
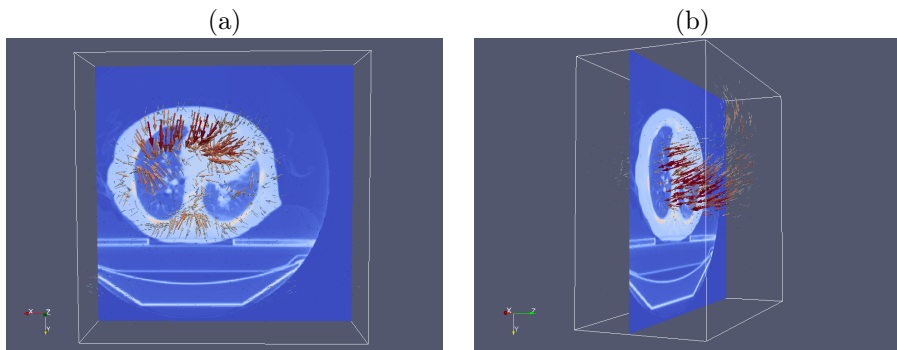

Figure 6.3: The (a) first and (b) second principal deformation basis functions analyzed from the RCCTs of a lung dataset. Colored lines indicate heat maps of the deformation magnitudes. As shown in the images, the first principal motion is the expansion / contraction of the lung and the second principal motion is along SI direction. $\overrightarrow{X}$: Left to Right (LR); $\overrightarrow{Y}$: Anterior to Posterior (AP); $\overrightarrow{Z}$: Superior to Inferior (SI).

To generate credible variations in the deformation space $\mathbf{C}$ for learning the linear operator $\mathbf{M}$, CLARET samples the largest scale of parameters by three standard deviations of the scores appeared in the planning image set $\{J_\tau$ over phase $\tau\}$ for each PCA-derived basis function. With the generated Fréchet mean image $\bar{J} = I$ the learning computed projections $\mathbf{P}(\mathbf{x}, I \circ T(\mathbf{C}); \theta)$ are generated by 1) transformation based on those credible variations $T(\mathbf{C})$ and 2) projection from a given tomographic angle θ to the transformed 3D volume $I \circ T(\mathbf{C})$.

In the registration, CLARET iteratively applies S multi-scale linear operators, from $\mathbf{M}_1$ to $\mathbf{M}_S$ to estimate the scores $\mathbf{C}$ on the basis functions ϕ_{pc} from current 2D intensity residues $\mathbf{R}$. The residues are formed by the difference between the normalized and histogram matched target projections $\Psi^\star$ (see figure 5.1 (c)) and the normalized projections (see figure 5.1 (d)) computed from the presently estimated deformation applied to the Fréchet mean image.

7. Results

We show CLARET's rigid registration and non-rigid registration results in sections 7.1 and 7.2, respectively. In particular, we tested the rigid registration

using the NST imaging system for the head-and-neck IGRT and tested the non-rigid registration using a CBCT imaging system for the lung IGRT.

7.1. Rigid Registration Results

We tested CLARET's rigid registration by synthetic treatment-time projections and by real phantom projections, as described in sections 7.1.1 and 7.1.2, respectively. The registration quality was measured by 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 at the center of the pharynx in the planning CT I.

$$mTRE(I) = \frac{1}{\chi} \sum_{i=1}^{\chi} \|I(\mathbf{y}_i) \circ T(\mathbf{C}_{true}) - I(\mathbf{y}_i) \circ T(\mathbf{C}_{est})\|_2 \qquad (7.1)$$

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 ith voxel position, and $\mathbf{C}_{true}$, $\mathbf{C}_{est}$ are the true and the estimated transformation parameters, respectively.

7.1.1. Synthetic Treatment Projections

We used noise-added DRRs (digitally-reconstructed projections) of target CTs as the synthetic treatment-time projections. The DRRs (see figure 7.1(a)) were generated to simulate the NST projections with dimension: 128×128; and pixel spacing: 3.2 mm (see figure 2.1). The target CTs were transformed from the patient's planning CT by taking normally distributed random samples of the translation and rotation parameters within ± 2 cm and $\pm 5°$. The planning CTs are with a voxel size = 1.2 mm lateral $\times$ 1.2 mm anterior-posterior $\times$ 3.0 mm superior-inferior. In the testing the number of imaging positions was varied to find the minimum number with sub-CT-voxel accuracy in terms of mTRE.

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 $\times$ (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 figure 7.1(b).

We 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 figure 7.2(a), show that two projection images are enough for CLARET to have sub-CT-voxel accuracy in the head and neck. Figure 7.1(c) shows the geometry of the two opposing x-ray sources that generated the two projection images in the study.

In addition, we studied the effect of the number of scales for CLARET learning. Figure 7.2(b) shows that increasing the number of scales for CLARET learning reduces the registration errors.
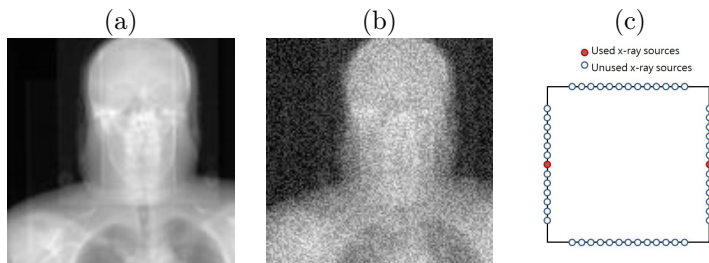
Figure 7.1: (a) A raw DRR from a x-ray source in the NST (b) DRR with Gaussian noise added (c) the NST geometry of two opposing x-ray sources
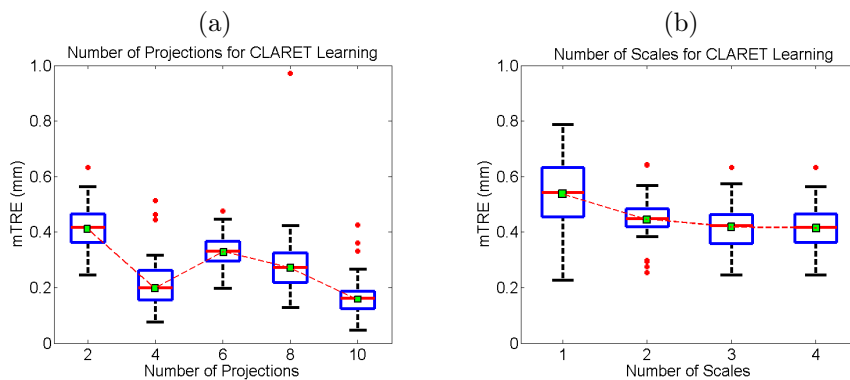

Figure 7.2: Boxplot results of errors in varying (a) the number of projections used and (b) the number of scales used for CLARET's rigid registration. Red dots are the outliers. In (a), projections of equally-spaced sources were used.

Table 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). In those test cases, CLARET performed registration using only the two opposing NST projection images. See the geometry in figure 7.1(c).

(mm; °)	Tx	Ty	Tz	Rx	Ry	Rz	mTRE
MAE	0.094	0.302	0.262	0.1489	0.0248	0.1540	0.524
SD	0.085	0.211	0.715	0.1093	0.0174	0.2824	0.728

Table 1: Mean absolute errors (MAE) and standard deviation (SD) of the absolute errors of the six rigid parameters appeared in the 90 synthetic test cases where CLARET used two synthetic NST projection images to do the registration.

7.1.2. Real Treatment Projections

We tested CLARET's rigid registration on a head-and-neck phantom dataset. NST projections (dimension: 1024×1024; pixel spacing: 0.4 mm) of the head-

and-neck phantom were downsampled to dimension 128×128 with a pixel spacing of 3.2 mm. The dimension of the planning CT is $512 \times 512 \times 96$ with a voxel size of 3.43 mm^3. The initial error between the planning time and the treatment time is 51.8 mm. With 4-scale learning ($S = 4$), CLARET obtained a sub-voxel accuracy of 3.32 mm using only two projections in 5.81 seconds. As shown in figure 7.3, CLARET is more accurate as more projections and scales used in its learning. The registration time is approximately linear to the number of projections used. The projection function $\mathbf{P}$ is implemented on a GPU (NVIDIA GeForce 9400m).
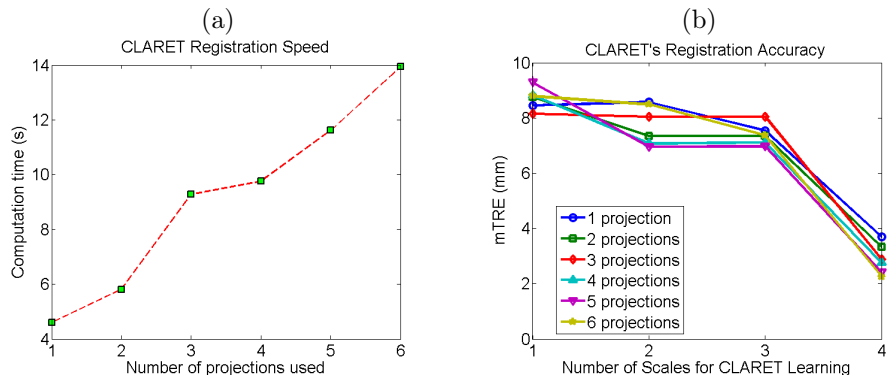

Figure 7.3: (a) Time plots and (b) error plots of CLARET's registrations on a real head-and-neck phantom dataset.

7.2. Non-rigid Registration Results

We tested CLARET's non-rigid registration with synthetic and real patient cone-beam projections, as described in sections 7.2.1 and 7.2.2, respectively. Respiratory-correlated CT ($RCCT$) datasets (CT dimension: $512 \times 512 \times 120$; voxel size = 1 mm lateral $\times$ 1 mm anterior-posterior $\times$ 2.5 mm superior-inferior) were generated by a 8-slice scanner (LightSpeed i, GE Medical Systems), acquiring repeat 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.

7.2.1. Synthetic Treatment Projections

We used DRRs of the target CTs as the synthetic treatment-time projections. The DRRs were generated to simulate projections in a gantry-mounted kV on-board imaging system (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 eigenmodes of the deformation of the patient's RCCTs.

For each one of the 10 CLARET's registrations, in our studies so far we used a *single* simulated coronal projection (dimension: 128×96; pixel spacing: 3.10 mm) at angle $= 14.18°$ (see figure 5.1(d)) as the input. (Future studies will investigate the effect of this pixel spacing on the registration quality.) 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 their active contour (Snake) segmentations (Yushkevich et al. [14]). As shown in Table 2, after registration CLARET can reduce more than 85% of centroid error.

Case#	1	2	3	4	5	6	7	8	9	10
Before	8.23	21.33	21.78	20.05	9.86	10.24	10.92	15.72	14.87	19.91
After	1.30	0.78	1.52	3.33	0.75	1.31	0.45	1.57	2.07	2.72

Table 2: 3D tumor centroid error (mm) before and after CLARET's registration for the 10 randomly generated test cases.

We studied CLARET's registration quality in average *DVF* (Displacement Vector Field) error over all cases and all CT voxels versus different angular spacings for learning. For each sampling angle, we generated 30 random test cases. Figure 7.4(a) shows the average DVF error reduces with appropriately large angular spacings.

We 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, we also generated 30 random test cases. Figure 7.4(b) shows no particular trend. As a result, we used a single projection to test CLARET's non-rigid registration for the real patient data in the next section.
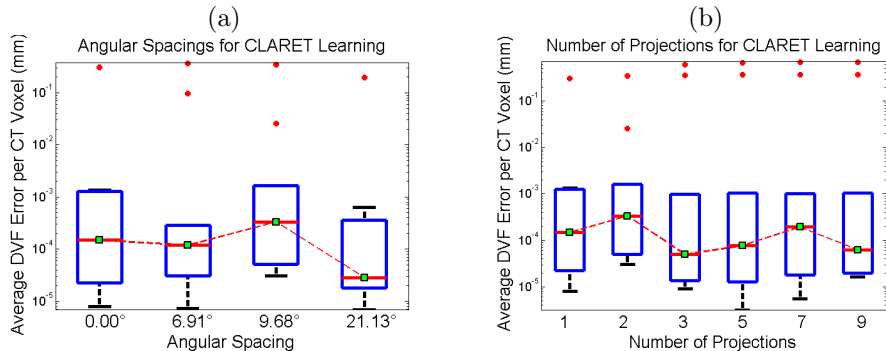

Figure 7.4: Boxplot results of errors in varying (a) the angular spacing and (b) the number of projections used for CLARET's non-rigid registration. Red dots are the outliers. In (a), two projections for each test were used. For the zero-degree test case, only one projection was used. In (b), DRRs spanning $9.68°$ about $14.18°$ were used in each test. For the single projection test case, it was tested at angle $= 14.18°$ (see figure 5.1(d)).

14

7.2.2. Real Treatment Projections

We tested CLARET on 5 lung dataset with the on-board CBCT system mentioned before. CBCT projections were shifted 16 cm to the left for acquiring CBCT scans in a half-fan mode. In this real patient study, a *single* coronal on-board CB projection (dimension: 1024×768; pixel spacing: 0.388 mm) down-sampled to dimension: 128×96 with pixel spacing: 3.10 mm at 14.18° (see figure 5.1(a)) at both the *EE* (End-Expiration) and the *EI* (End-Inspiration) phases were used for CLARET's testing. We measured the 3D tumor centroid (see figure 7.6(a)) difference between CLARET-estimated CT and the reconstructed CBCT at the same respiratory phase as the testing projection. For the Gaussian normalization, we set the RMS width of the Gaussian window to 31.98 mm for this imaging geometry where CLARET yielded the smallest 3D centroid error for a lung dataset (see figure 7.5). (Future studies will check whether this window size is also best for other datasets.) The results shown in Table 3 suggest a consistency of registration quality between the synthetic tests and the real patient tests. The average computation time is 5 seconds with the projection function **P** implemented on a GPU (NVIDIA GeForce 9800 GTX).

CLARET reduces errors in the directions orthogonal to the projection direction (coronal plane) more than those in the projection direction. As shown in Table 3, most of the percentages of the 2D error reduction (coronal plane error reduction), except cases from patient #1, are greater than those of the 3D error reduction. This is expected because the single projection used communicates tumor positions directly in the two dimensions orthogonal to the projection directions but communicates tumor positions along the projection direction only from perspective effects. For directions orthogonal to the projection directions, CLARET can particularly accurately locate the tumor. See figure 7.7 discussed in the next section.

Patient#	1	2	3	4	5
EE, 3D, before	7.96	9.70	1.47	10.17	3.52
EE, 3D, after	2.27 (72%)	3.20 (67%)	1.32 (10%)	2.77 (73%)	2.24 (36%)
EE, 2D, before	4.16	9.18	1.47	9.67	3.46
EE, 2D, after	2.07 (50%)	2.23 (76%)	1.32 (10%)	1.51 (84%)	1.91 (45%)
EI, 3D, before	8.03	7.45	3.63	5.53	3.89
EI, 3D, after	5.26 (34%)	2.85 (62%)	2.03 (44%)	2.31 (58%)	2.40 (38%)
EI, 2D, before	6.80	6.77	3.49	5.17	3.16
EI, 2D, after	4.71 (31%)	1.64 (75%)	1.86 (47%)	1.52 (71%)	0.83 (74%)

Table 3: 2D and 3D tumor centroid errors (mm) of EE and EI phases before and after CLARET's registration for 5 patients' on-board CBCT dataset. The number inside the parentheses indicates the percentage of the error reduction. The 2D error refers to the tumor centroid error in the directions orthogonal to the projection direction (coronal plane).
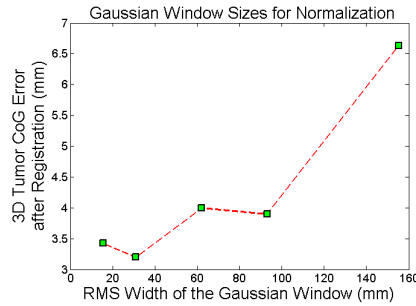
Figure 7.5: 3D tumor centroid error plots on a lung dataset with varying sizes of the Gaussian window used for CLARET's local Gaussian normalization.

7.2.3. Visual Validation

Figure 7.6(b) shows the 3D meshes of the tumors in the mean CT, the CBCT at EE respiratory phase, and the estimated CT of a lung dataset for visual validation. As shown in the figure, CLARET moves the tumor up in the lung from the mean image; this is expected physiologically for the EE phase. Figure 7.7 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 fact that the lines pass near the tumor centroid in the CLARET-estimated CT and results shown in Table 3 indicate that CLARET can accurately locate the tumor in the directions orthogonal to the projection direction (coronal plane).
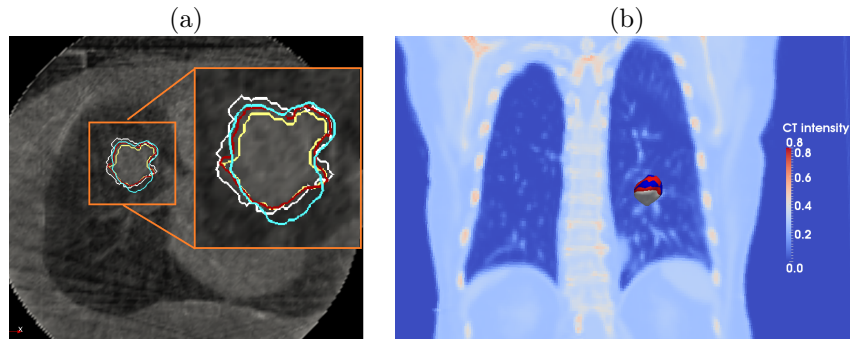

Figure 7.6: (a) Manual segmented contours in the reconstructed CBCT at specific phase. They were used for 3D centroid calculation. (b) Tumor meshes in the mean CT (gray), in the target CBCT at the EE respiratory phase (blue) and in the CLARET-estimated CT (red) of a lung dataset. The background is a coronal slice of the mean CT for illustration. The overlap between the estimated and the target tumor meshes indicates a good registration.
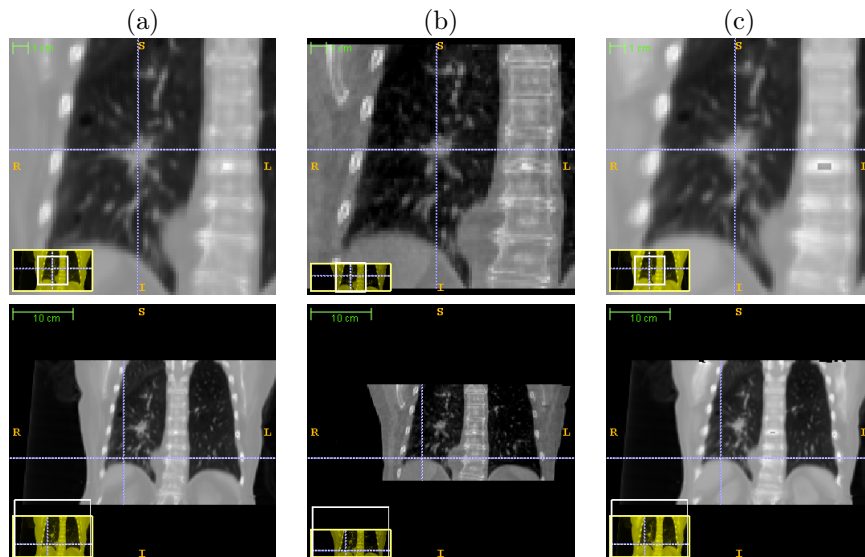
16

Figure 7.7: 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 7.6(b). Upper row: lines locating the tumor centroid in the CBCT at the EE phase; lower row: lines locating the diaphragm contour in the CBCT at the EE phase.

8. Conclusions and Discussion

We presented a novel rigid and non-rigid 2D/3D registration method that estimates an image region's 3D motion/deformation parameters from a very small set of 2D projection images of that region. The method is based on producing limited-dimension parameterizations of geometric transformations based on the region's 3D images. The method operates via iterative, multi-scale regression, where the regression matrices are learned in a way specific to the 3D image(s). Relative to registration methods based on optimization, the method is more robust, has wider capture range, and is qualitatively faster while achieving comparable registration accuracy. The synthetic and real test results have shown our method's promise to provide fast and accurate tumor localization with a small set of treatment-time imaging projections for IGRT.

However, in order to obtain such robustness, our registration method requires a well-modeled motion/deformation shape space that includes all credible variations of the image region. In many radiation therapy situations for certain parts of the body, collecting the required number of 3D images of the patient to form the well-modeled shape space is not directly obtainable in current therapeutic practice.

To make our method more robust for IGRT application, the future work of this paper is to evaluate the method on more patient datasets and to study the effects of the projection resolution and the normalization window size on the

registration accuracy.

9. Acknowledgement

References

[1] D. Russakoff, T. Rohlfing, C. Maurer, Fast intensity-based 2D-3D image registration of clinical data using light fields, in: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, vol. 1, 416–422, 2003.

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

[3] A. Khamene, P. Bloch, W. Wein, M. Svatos, F. Sauer, Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy, Medical Image Analysis 10 (2006) 96–112.

[4] R. Munbodh, D. A. Jaffray, D. J. Moseley, Z. Chen, J. P. S. Knisely, P. Cathier, J. S. Duncan, Automated 2D-3D registration of a radiograph and a cone beam CT using line-segment enhancement, Medical Physics 33 (2006) 1398–1411.

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

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

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

[8] F. Jurie, M. Dhome, Hyperplane Approximation for Template Matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 996–1000.

[9] M. F. Beg, M. I. Miller, A. Trouve, L. Younes, Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms, International Journal of Computer Vision 61 (2005) 139–157.

[10] G. E. Christensen, R. D. Rabbitt, M. I. Miller, Deformable Templates Using Large Deformation Kinematics, IEEE Transactions on Image Processing 5 (1996) 1435–1447.

[11] B. C. Davis, E. Bullitt, P. T. Fletcher, S. Joshi, S.: Population shape regression from random design data, in: In: Proc. of ICCV 2007, 1–7, 2007.

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

[13] X. Liu, B. Davis, M. Niethammer, S. Pizer, G. Mageras., Prediction-driven Respiratory Motion Atlas Formation for 4D Image- guided Radiation Therapy in Lung, in: MICCAI'10 Pulmonary Image Analysis Workshop, 2010.

[14] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, G. Gerig, User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability, Neuroimage 31 (3) (2006) 1116–1128.