Tissue-Volume Preserving Deformable Image Registration for 4DCT Pulmonary Images

Bowen Zhao, Gary E. Christensen, Joo Hyun Song, Yue Pan, Sarah E. Gerard, Joseph M. Reinhardt
University of Iowa, Iowa City, Iowa 52242
{bowen-zhao, gary-christensen, joohyun-song, yue-pan-1, sarah-gerard, joe-reinhardt}@uiowa.edu

Kaifang Du, Taylor Patton, John M. Bayouth
University of Wisconsin-Madison
Madison, WI 53706
{kaifangdu, bayouth}@humonc.wisc.edu
tpatton2@wisc.edu

Geoffrey D. Hugo
Virginia Commonwealth University
401 College St. Richmond, VA 23229
gdhugo@vcu.edu

Abstract

We propose a 4D (three spatial dimensions plus time) tissue-volume preserving non-rigid image registration algorithm for pulmonary 4D computed tomography (4DCT) data sets to provide relevant information for radiation therapy and estimate pulmonary ventilation. The sum of squared tissue volume difference (SSTVD) similarity cost takes into account the CT intensity changes of spatially corresponding voxels, which is caused by the variations of fraction of tissue within voxels throughout the respiratory cycle. The proposed 4D SSTVD registration scheme considers the entire dynamic 4D data set simultaneously, using both spatial and temporal information. We employed a uniform 4D cubic B-spline parametrization of the transform and a temporally extended linear elasticity regularization of deformation field to ensure temporal smoothness and thus biological plausibility of estimated deformation. We used a multi-resolution multi-grid registration framework with limited-memory Broyden Fletcher Goldfarb Shanno (L-BFGS) optimization procedure for rapid convergence and limited memory consumption. We conducted experiments using synthetic 2D+t images and clinical 4DCT pulmonary data sets and evaluated accuracy and temporal smoothness of the proposed method via manually annotated landmarks.

Keywords: *4D image registration, groupwise image registration, sum of squared tissue volume difference, mass preservation, temporal smoothness, radiation therapy, treatment planning*

1. Introduction

In pulmonary radiation therapy, target localization and motion tracking of parenchyma, as well as tumor regions, are critically important for guiding radiation on pathology, minimizing dose for normal tissue and estimating regional lung ventilation[11]. Time-resolved dynamic imaging data sets have become increasingly available with the advancement of medical imaging techniques[7], making it feasible to track and model pulmonary motion. Pulmonary 4DCT allows for reconstruction of multiple 3D volumes corresponding to different breathing phases throughout the respiratory cycle. Non-rigid image registration is used to estimate and track correspondences between these phase images.

The CT number of lung parenchyma varies across the breathing phases due to changes in fraction of tissue within voxels, which is caused by air flowing in and out of the lung throughout the respiratory cycle. Non-rigid registration of 4DCT pulmonary images is a difficult problem due to the fact that spatially corresponding points belonging to different breathing phases have dissimilar image intensities because of lung ventilation. To compensate for these intensity variations, and to enforce the principle of tissue conservation, Yin et al.[18] and Gorbunova et al.[5] independently proposed an intensity-based registration similarity cost function that accounts for these intensity changes across breathing phases. We will refer to this previously established cost function as the sum of squared tissue volume differences (SSTVD). The SSTVD cost function achieves desirable registration results for pulmonary CT image alignment compared to the conventional sum of squared (intensity) difference (SSD) cost function [18],[5] which does not account for intensity changes. In this paper, we extended the SSTVD registration framework from 3D to 4D.

4D registration techniques have become more common in recent years[10][17]. Unlike traditional registration schemes which perform a sequence of 3D pair-wise registrations, either with respect to a common reference or between consecutive time point images, 4D registration frameworks consider the information contained within the entire time-resolved imaging data set simultaneously without bias toward any chosen reference image. One advantage of 4D image registration is that it avoids accumulation of

numerical and discretization errors associated with 3D sequential registration. Additionally, 4D registration frameworks provide smooth and consistent displacement fields along spatial and temporal dimensions[14]. Thus, 4D registration would provide more biologically reasonable tissue motion tracking than 3D pair-wise registration. This could be used to improve radiation therapy plans and dose delivery during breathing [9].

In this paper, we propose a 4D tissue conservation registration algorithm (4D SSTVD) which incorporates the merits of SSTVD registration similarity cost with a 4D transformation model. The resulting registration framework is more suitable for 4DCT pulmonary image registration than pairwise 3D SSTVD and 4D SSD [10] registration.

2. Methods

Our proposed 4D SSTVD image registration method minimizes an objective function consisting of a 4D SSTVD similarity cost and a temporally extended linear elastic regularization cost. In this paper, the concept of *time* is identified with that of *breathing phases* for the 4DCT image, indicating that the 3D image at *time point t* is the same as the *phase t* 3D image. Also, the term fixed image will be used interchangeably with target image in the registration framework.

2.1. Review of SSTVD Similarity Cost Function

The underlying assumption behind SSTVD similarity cost is that the total tissue volume will remain roughly the same throughout the respiratory cycle while the total lung volume will change as air flows in and out of the lung. Let $I_f^{HU} : \Omega_f \to \mathbb{R}, \mathbf{x} \mapsto I_f^{HU}(\mathbf{x})$ denote the fixed image in Hounsfield units and $I_m^{HU} : \Omega_m \to \mathbb{R}, \mathbf{y} \mapsto I_m^{HU}(\mathbf{y})$ the moving image in Hounsfield units. Following the rationale presented in the work of Yin et al.[18], an intensity linear transformation is performed on these images and consequently the intensity value at each voxel is converted from radiodensity in Hounsfield units to actual tissue volume in mm^3. Specifically, we have:

$$\begin{aligned} I_f(\mathbf{x}) &= v_f \cdot r_f(\mathbf{x}) = v_0 \cdot r_f(\mathbf{x}) \\ I_m(\mathbf{y}) &= v_m \cdot r_m(\mathbf{y}) = v_0 \cdot r_m(\mathbf{y}) \end{aligned} \tag{1}$$

where $r_f(\mathbf{x})$ and $r_m(\mathbf{y})$ represent the fraction of tissue within a standard image voxel centered at $\mathbf{x}$ and $\mathbf{y}$ for the fixed and moving image, respectively, and they are given by

$$\begin{aligned} r_f(\mathbf{x}) &= \frac{I_f^{HU}(\mathbf{x}) - HU_{air}}{HU_{tissue} - HU_{air}} = \frac{I_f^{HU}(\mathbf{x}) + 1000}{1055} \\ r_m(\mathbf{y}) &= \frac{I_m^{HU}(\mathbf{y}) - HU_{air}}{HU_{tissue} - HU_{air}} = \frac{I_m^{HU}(\mathbf{y}) + 1000}{1055} \end{aligned} \tag{2}$$

where $HU_{tissue} = 55$ and $HU_{air} = -1000$ are the radiodensity values of tissue and air in Hounsfield units. v_f and v_m are the volumes of a standard voxel in the fixed and moving image, respectively. For images in the same 4DCT data set, we usually have $v_f = v_m = v_0$, i.e., standard voxels have constant volume in both the fixed and moving image, which can be calculated from physical spacing of the data set. The resulting images $I_f : \Omega_f \to \mathbb{R}$ and $I_m : \Omega_m \to \mathbb{R}$ are named tissue-volume images and it is these two images that we are trying to register in the SSTVD registration framework.

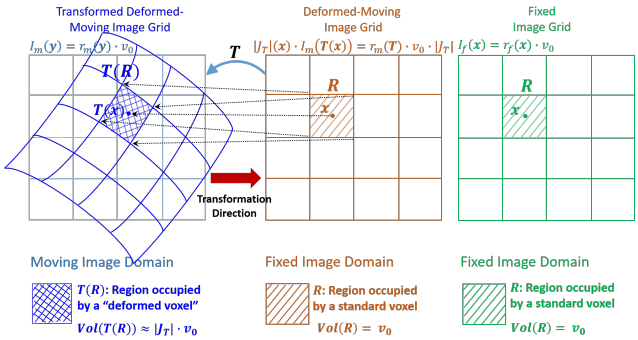


Figure 1: Illustration of SSTVD cost and the "deformed voxel".

Let $\mathbf{T} : \Omega_f \to \Omega_m, \mathbf{x} \mapsto \mathbf{T}(\mathbf{x})$ be the geometric transform deforming I_m to match I_f. The SSTVD similarity cost is thus given by:

$$\begin{aligned} & C(\mathbf{T}) \\ &= \frac{1}{|\Omega_f|} \sum_{\mathbf{x} \in \Omega_f} \left(v_0 \cdot r_f(\mathbf{x}) - (|J_\mathbf{T}|(\mathbf{x}) \cdot v_0) \cdot r_m(\mathbf{T}(\mathbf{x})) \right)^2 \\ &= \frac{1}{|\Omega_f|} \sum_{\mathbf{x} \in \Omega_f} \left(I_f(\mathbf{x}) - |J_\mathbf{T}|(\mathbf{x}) \cdot I_m(\mathbf{T}(\mathbf{x})) \right)^2 \end{aligned} \tag{3}$$

where $v_0 \cdot r_f(\mathbf{x})$ is the volume of tissue within a standard voxel centered at $\mathbf{x}$ in the fixed tissue-volume image I_f, while $|J_\mathbf{T}|(\mathbf{x}) \cdot v_0 \cdot r_m(\mathbf{T}(\mathbf{x}))$ is the volume of tissue within a standard voxel centered at $\mathbf{x}$ in the deformed moving tissue-volume image. The spatial Jacobian of transform $|J_\mathbf{T}|(\mathbf{x})$ is introduced because the mathematical interpretation of transform Jacobian $|J_\mathbf{T}|(\mathbf{x})$ is the contraction or expansion of an infinitesimal local neighborhood of $\mathbf{x}$ induced by $\mathbf{T}$. Let $\mathbf{R}$ denote the region occupied by the standard voxel centered at $\mathbf{x}$ in the fixed image domain, whose volume is v_0. Then $\mathbf{T}(\mathbf{R})(\ni \mathbf{T}(\mathbf{x}))$ would be the transformed region occupied by the "deformed voxel" in the moving image domain, whose volume can be approximated by $|J_\mathbf{T}|(\mathbf{x}) \cdot v_0$, as illustrated in Figure 1. In the SSTVD registration scheme, the term $|J_\mathbf{T}|(\mathbf{x}) \cdot I_m(\mathbf{T}(\mathbf{x}))$ as a whole serves as the deformed moving image, instead of $I_m(\mathbf{T}(\mathbf{x}))$ alone.

Compared to conventional SSD cost function, which is

given as below (by abuse of notation for I_f and I_m):

$$C_{SSD}(\mathbf{T}) = \frac{1}{|\Omega_f|} \sum_{\mathbf{x}\in\Omega_f} \Big(I_f(\mathbf{x}) - I_m(\mathbf{T}(\mathbf{x})) \Big)^2 \quad (4)$$

the additional Jacobian modification $|J_\mathbf{T}|(\mathbf{x})$ will help compensate for the intensity changes of spatially corresponding voxels in a 4DCT image data set, which is caused by periodic variations in the fraction of tissue within voxels during ventilation.

2.2. 4D SSTVD Similarity Cost Function

Inspired by the work of Yin et al.[18], Gorbunova et al.[5] and Metz et al.[10], the proposed 4D SSTVD similarity cost function takes advantage of the fact that the tissue volume can be roughly considered as constant [18] throughout the respiratory cycle and it uses the spatial and temporal information within the 4D data set simultaneously.

Before proceeding to the actual formulation of 4D SSTVD cost, we shall define some notations. Denote the generic spatio-temporal coordinate as $\mathbf{x} = (x_1, x_2, x_3, x_4)^T \equiv (x, y, z, t)^T$. Define the 4DCT image domain as $\Omega = \Omega_s \times \Omega_t$, where $\Omega_s = \Omega_x \times \Omega_y \times \Omega_z$ is the spatial domain and $\Omega_t = \{0, 1, ..., |\Omega_t| - 1\}$ is the temporal domain. Also, let $\mathbf{x}_s = (x, y, z)^T \in \Omega_s$ be the vector of spatial coordinates, $t \in \Omega_t$ the time/phase coordinate, and thus $\mathbf{x} = (\mathbf{x}_s^T, t)^T \in \Omega$ represents the generic spatio-temporal coordinates of a point in the 4DCT image domain. Denote $\mathbf{T} : \Omega_s \times \Omega_t \to \Omega_s \times \Omega_t$ as the transformation from fixed/target image domain to moving image domain with the constraint that $\mathbf{T}(\mathbf{x}_s, t) = (\phi(\mathbf{x}_s)^T, t)^T$ where $\phi : \Omega_s \to \Omega_s$, i.e., no temporal displacement should occur. This is to ensure that any voxel within a certain phase image will remain in that phase without being temporally deformed and moving forward or backward in time. Finally, let $|J_\mathbf{T}|(\mathbf{x})$ be the local Jacobian of transformation $\mathbf{T}$ evaluated at $\mathbf{x}$.

The original 4DCT image $I^{HU} : \Omega \to \mathbb{R}$ underwent the intensity affine transformation in Equ.5 and was converted into 4D tissue-volume image $I : \Omega \to \mathbb{R}$. Note, the obtained tissue-fraction image $r(\mathbf{x}) = (I^{HU}(\mathbf{x}) + 1000)/1055$ was windowed to fall in range $[0, 1]$, and the scaling factor v_0 was dropped in actual implementation.

$$I(\mathbf{x}) = v_0 \cdot \frac{I^{HU}(\mathbf{x}) - HU_{air}}{HU_{tissue} - HU_{air}} = v_0 \cdot \frac{I^{HU}(\mathbf{x}) + 1000}{1055} \quad (5)$$

The proposed 4D SSTVD intensity cost function uses the 4D tissue-volume image as the moving image and temporal average of deformed moving tissue-volume image as the target image. Under this construction, the target image is given implicitly and will be iteratively updated with the optimization process until it converges to a relatively stable state. Note that at the beginning of the algorithm, the transformation is identity and the transform Jacobian is 1 everywhere, so the initial target image is just the temporal average of the entire 4D tissue-volume image.

$$C_{int}(\mathbf{T}) = \frac{1}{|\Omega_s|} \frac{1}{|\Omega_t|} \sum_{\mathbf{x}_s\in\Omega_s} \sum_{t\in\Omega_t} \Big(\bar{K}(\mathbf{x}_s) - K(\mathbf{x}) \Big)^2 \quad (6)$$

where $K(\mathbf{x}) = |J_\mathbf{T}|(\mathbf{x}) \cdot I(\mathbf{T}(\mathbf{x}))$ is the deformed moving tissue-volume image, and

$$\begin{aligned} \bar{K}(\mathbf{x}_s) &= \frac{1}{|\Omega_t|} \sum_{\tau\in\Omega_t} K(\mathbf{x}_s, \tau) \\ &= \frac{1}{|\Omega_t|} \sum_{\tau\in\Omega_t} \Big(|J_\mathbf{T}|(\mathbf{x}_s, \tau) \cdot I(\mathbf{T}(\mathbf{x}_s, \tau)) \Big) \end{aligned} \quad (7)$$

is temporal average of deformed moving tissue-volume image and serves as target image in the registration framework.

2.3. 4D Transformation Model

The 4D transformation $\mathbf{T}$ is parameterized using a tensor product of four 1D cubic B-spline[12] kernels and is given by Equation 8. The notation $\mathbf{T}(\mathbf{x}, \mathbf{a})$ is used to emphasize the dependence of transformation $\mathbf{T}$ on spatio-temporal coordinates $\mathbf{x}$ as well as the parameters $\mathbf{a}$ used to parameterize the transformation.

$$\begin{aligned} \mathbf{T}(\mathbf{x}, \mathbf{a}) &= \mathbf{x} + \mathbf{u}(\mathbf{x}, \mathbf{a}) \\ &= \mathbf{x} + \sum_{i=-1}^{N_x-2} \sum_{j=-1}^{N_y-2} \sum_{k=-1}^{N_z-2} \sum_{l=-1}^{N_t-2} \mathbf{a}_{i,j,k,l} \\ &\quad B(\frac{x}{\delta_x} - i) B(\frac{y}{\delta_y} - j) B(\frac{z}{\delta_z} - k) B(\frac{t}{\delta_t} - l) \end{aligned} \quad (8)$$

where $\mathbf{u}(\mathbf{x}, \mathbf{a})$ is the displacement vector dependent on spatio-temporal coordinates $\mathbf{x}$ and B-spline coefficients $\mathbf{a}$. $\mathbf{a}_{i,j,k,l} \in \mathbb{R}^3 \times \{0\}$ is the B-spline coefficient vector at B-spline grid point $(i, j, k, l) \in \mathbb{Z}^4$. Note that setting the last coordinate of all four dimensional B-spline coefficient vectors $\mathbf{a}_{i,j,k,l}$ to 0 effectively prevents any temporal transformation. The variables N_α and δ_α for $\alpha \in \{x, y, z, t\}$ represent the number of grid points and grid spacing in the x, y, z and t directions, respectively.

Even though temporal displacement was restricted to 0, the 4D B-spline parametrization was used to ensure the smoothness of all spatial displacement components along the temporal direction.

2.4. Temporally Extended Linear Elasticity Regularization

We used a temporally extended linear elastic model to regularize the 4D SSTVD image registration algorithm. The conventional spatial linear elasticity regularization is valid

for small elastic deformations between a pair of phase images within the breathing cycle[3]. Based on the spatial linear elastic model, we imposed an additional temporal smoothness regularization on spatial displacement components and formulated the temporally extended linear elasticity regularization cost as in Equ. 9:

$$C_{reg}(\mathbf{a}) = \sum_{\mathbf{x}\in\Omega} ||(L\mathbf{u})(\mathbf{x},\mathbf{a})||^2 = \sum_{\mathbf{x}\in\Omega}\sum_{i=1}^{4} L_i^2(\mathbf{x},\mathbf{a}) \quad (9)$$

where the linear elasticity differential operator L is given by (omitting arguments $\mathbf{x}$ and $\mathbf{a}$)

$$\begin{aligned} L\mathbf{u} &= c_1(\nabla\cdot\nabla)\mathbf{u} + c_2\nabla(\nabla\cdot\mathbf{u}) + c_3\mathbf{u} \\ &= (L_1, L_2, L_3, L_4)^T \end{aligned} \quad (10)$$

where $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z, \partial/\partial t)^T$ is the gradient operator. Examples of expressions for L_i are given by:

$$\begin{aligned} L_1 = c_1(&\frac{\partial^2 u_x}{\partial x^2} + \frac{\partial^2 u_x}{\partial y^2} + \frac{\partial^2 u_x}{\partial z^2} + \frac{\partial^2 u_x}{\partial t^2}) \\ &+ c_2(\frac{\partial^2 u_x}{\partial x^2} + \frac{\partial^2 u_y}{\partial x\partial y} + \frac{\partial^2 u_z}{\partial x\partial z}) + c_3 u_x \end{aligned} \quad (11)$$

$$L_4 = c_2(\frac{\partial^2 u_x}{\partial t\partial x} + \frac{\partial^2 u_y}{\partial t\partial y} + \frac{\partial^2 u_z}{\partial t\partial z}) \quad (12)$$

The constants c_1, c_2 and c_3 are the weights that adjust the elasticity of the model. In our experiments, we chose $c_1 = 0.75$, $c_2 = 0.25$ and $c_3 = 0$.

2.5. Optimization Strategy

The 4D SSTVD registration algorithm is implemented by finding the transformation parameters that minimize a linear combination of the intensity similarity cost and the linear elasticity penalty term

$$C_{total}(\mathbf{a}) = \lambda C_{int}(\mathbf{a}) + C_{reg}(\mathbf{a}) \quad (13)$$

where λ is a weight to be tuned. For the experiments in this paper, we chose $\lambda = 80$.

Minimizing the cost function necessitates acquiring its derivatives with respect to the transform parameters. Let $b_\alpha \triangleq a_{p,q,r,s,\alpha}, \alpha \in \{x, y, z\}$, be the α component of the B-spline coefficient vector $\mathbf{a}_{p,q,r,s}$ at grid location $[p, q, r, s] \in \mathbb{Z}^4$. Using the chain rule, product rule, and formulas from matrix calculus, the derivative of 4D SSTVD intensity cost is given by

$$\begin{aligned} &\frac{\partial C_{int}(\mathbf{a})}{\partial b_\alpha} \\ &= \frac{\partial}{\partial b_\alpha}\frac{1}{|\Omega_s|}\frac{1}{|\Omega_t|}\sum_{\mathbf{x}_s\in\Omega_s}\sum_{t\in\Omega_t}\left(\bar{K}(\mathbf{x}_s,\mathbf{a}) - K(\mathbf{x},\mathbf{a})\right)^2 \\ &= \frac{2}{|\Omega_s|}\sum_{\mathbf{x}_s\in\Omega_s}\frac{1}{|\Omega_t|}\sum_{t\in\Omega_t}\left[\left(\bar{K}(\mathbf{x}_s,\mathbf{a}) - K(\mathbf{x},\mathbf{a})\right)\cdot \right. \\ &\quad \left.\left(\frac{\partial}{\partial b_\alpha}\bar{K}(\mathbf{x}_s,\mathbf{a}) - \frac{\partial}{\partial b_\alpha}K(\mathbf{x},\mathbf{a})\right)\right] \\ &= -\frac{2}{|\Omega_s|}\sum_{\mathbf{x}_s\in\Omega_s}\frac{1}{|\Omega_t|}\sum_{t\in\Omega_t}\left[\left(\bar{K}(\mathbf{x}_s,\mathbf{a}) - K(\mathbf{x},\mathbf{a})\right)\cdot \right. \\ &\quad \left.\frac{\partial}{\partial b_\alpha}K(\mathbf{x},\mathbf{a})\right] \end{aligned} \quad (14)$$

where (omitting parameters $\mathbf{x}$ and $\mathbf{a}$),

$$\begin{aligned} &\frac{\partial}{\partial b_\alpha}K(\mathbf{x},\mathbf{a}) \\ &= \frac{\partial|J_T|}{\partial b_\alpha}\cdot I(\mathbf{T}) + |J_T|\cdot\frac{\partial I(\mathbf{T})}{\partial b_\alpha} \\ &=< \frac{\partial|J_T|}{\partial J_T}, \frac{\partial J_T}{\partial b_\alpha} > \cdot I(\mathbf{T}) + |J_T|\cdot\nabla I(\mathbf{T})^T\frac{\partial\mathbf{T}}{\partial b_\alpha} \\ &=< cof J_T, \frac{\partial J_T}{\partial b_\alpha} > \cdot I(\mathbf{T}) + |J_T|\cdot\nabla I(\mathbf{T})^T\frac{\partial\mathbf{T}}{\partial b_\alpha} \\ &= |J_T|\left(< J_T^{-T}, \frac{\partial J_T}{\partial b_\alpha} > \cdot I(\mathbf{T}) + \nabla I(\mathbf{T})^T\frac{\partial\mathbf{T}}{\partial b_\alpha}\right) \end{aligned} \quad (15)$$

And derivative of the linear elasticity term is:

$$\frac{\partial C_{reg}(\mathbf{a})}{\partial b_\alpha} = \frac{\partial}{\partial b_\alpha}\sum_{\mathbf{x}\in\Omega}||L\mathbf{u}||^2 = 2\sum_{\mathbf{x}\in\Omega}\sum_{i=1}^{4}\left(L_i\cdot\frac{\partial L_i}{\partial b_\alpha}\right) \quad (16)$$

Image registration is often an ill-posed problem with many locally optimal solutions. Without the presence of an explicit target image, 4D registration is even worse in this regard. For example, since the temporal average of the deformed moving tissue-volume images is used as the registration target, two transformations differing by any rigid motions will result in the same value of intensity based cost. Therefore, similar to the work of Balci et al.[1] and Metz et al. [10], we impose a further constraint that the temporal average of the displacement field at any spatial location should be zero. Specifically, after every iteration, the derivative with respect to every B-spline coefficient will be modified as below, so that the derivatives at any spatial grid point will sum up to zero across time, resulting in zero average spatial displacement at any spatial location: $\forall\,[p, q, r, s] \in \mathbb{Z}^4$,

$$\frac{\partial C}{\partial a_{p,q,r,s,\alpha}} \leftarrow \frac{\partial C}{\partial a_{p,q,r,s,\alpha}} - \frac{1}{|\Omega_t|}\sum_{\tau\in\Omega_t}\frac{\partial C}{\partial a_{p,q,r,\tau,\alpha}} \quad (17)$$

We used an L-BFGS [8] optimizer in the registration framework because of its rapid quadratic convergence rate and limited memory consumption.

2.6. Inverse Transform

In order to get pair-wise spatial transforms between any two phase images, it is necessary to estimate the inverse transform $\mathbf{S} : \Omega \to \Omega$ that deforms the temporal average of deformed moving tissue-volume image (the target image) to match the original 4D tissue-volume image (the moving image). Because the B-spline parameterization does not have a closed form inverse expression, we follow the approach in [10] and estimate the inverse transform separately using a finer B-spline grid[10]. We minimize the following distance cost[2] to find the inverse transformation

$$C_{inv}(\mathbf{a}') = \sum_{\mathbf{x}\in\Omega} ||\mathbf{S}(\mathbf{T}(\mathbf{x}), \mathbf{a}') - \mathbf{x}||^2 \tag{18}$$

where $\mathbf{a}'$ is the set of B-spline coefficients parameterizing the inverse 4D transform $\mathbf{S}$. After the optimal forward and inverse transforms $\hat{\mathbf{T}}$ and $\hat{\mathbf{S}}$ have been estimated, we can compose them to acquire spatial transforms between any pair of two time point images $\mu, \nu \in \Omega_t$.

$$\mathbf{T}_{\nu\to\mu}(\mathbf{y}_s) = (\hat{\mathbf{T}}_\mu \circ \hat{\mathbf{S}}_\nu)(\mathbf{y}_s) \tag{19}$$

where $\mathbf{y}_s = (y_1, y_2, y_3)^T$ is a spatial point in time point image ν. $\hat{\mathbf{T}}_\mu$ is the spatial transform acquired by evaluating the optimal 4D forward transform at time point μ and $\hat{\mathbf{S}}_\nu$ is the spatial transform acquired by evaluating the optimal 4D inverse transform at time point ν.

2.7. Implementation

The proposed 4D SSTVD algorithm was implemented using Elastix package[6] based on Insight Segmentation and Registration Toolkit (ITK) libraries. This algorithm was implemented with multi-threading support. The experiments using clinical 4DCT data sets were run on a machine with Intel i7 5930k CPU (6 cores, 12 threads @ 3.5GHz) and 64GB of DDR4 memory. For each 4DCT data set, we used a multi-grid multi-resolution scheme consisting of 8 resolutions from coarse to fine to estimate both the forward and inverse transforms. The whole process took approximately 2 to 3 hours for each 4DCT data set.

2.8. Evaluation Methods

Evaluation of 4D registration frameworks needs to consider both accuracy and temporal smoothness [10][1][17]. 4D registration may not necessarily yield the best accuracy of registration results, but it generates deformation fields that are a lot more temporally smooth and consistent. With comparable or slightly worse accuracy, a temporally smoother outcome is preferred because the resulting estimated motion is biologically more reasonable.

For registration accuracy evaluation, we used the Mean Landmark Error (MLE)[13]. MLE measures the average distance between expert-labeled landmark positions on all phase images (except for the one chosen as reference) and landmark positions obtained by transforming the landmarks from a chosen reference phase onto all other phases. Let $\mathbf{p}_{r,i}$ be the ith landmark on the chosen reference time point r and $\mathbf{T}_{r\to t}$ be the spatial transform from time point r to t. Then $\mathbf{T}_{r\to t}(\mathbf{p}_{r,i}; \mathbf{a})$ is the transformed location in time point t of the landmark $\mathbf{p}_{r,i}$, while $\mathbf{p}_{t,i}$ is the real location of the corresponding landmark in time point t. The Mean Landmark Error is formulated as in Equation 20.

$$MLE(\mathbf{a}) = \frac{1}{N(|\Omega_t| - 1)} \sum_{\Omega_t \ni t \neq r} \sum_{i\in N} ||\mathbf{T}_{r\to t}(\mathbf{p}_{r,i}; \mathbf{a}) - \mathbf{p}_{t,i}|| \tag{20}$$

where N is the number of landmarks on every time point image.

For temporal smoothness evaluation, we used the Mean Irregularity of estimated landmark trajectories:

$$MIR(\mathbf{a}) = \frac{1}{N|\Omega_t|} \sum_{t\in\Omega_t} \sum_{i\in N} ||\frac{\partial^2 \mathbf{T}_{r\to t}(\mathbf{p}_{r,i}; \mathbf{a})}{\partial t^2}|| \tag{21}$$

where the second order partial derivatives are computed using centered finite difference assuming Neumann boundary conditions. Denote $\mathbf{q}_{t,i} = \mathbf{T}_{r\to t}(\mathbf{p}_{r,i}; \mathbf{a})$ and we have:

$$\begin{aligned} ||\frac{\partial^2 \mathbf{T}_{r\to t}(\mathbf{p}_{r,i}; \mathbf{a})}{\partial t^2}|| &= \sqrt{\sum_{\alpha\in\{x,y,z\}} \left(\frac{\partial^2 q_{t,i,\alpha}}{\partial t^2}\right)^2} \\ &= \sqrt{\sum_{\alpha} \left(\frac{q_{t+1,i,\alpha} - 2q_{t,i,\alpha} + q_{t-1,i,\alpha}}{(\Delta t)^2}\right)^2} + O((\Delta t)^2) \end{aligned} \tag{22}$$

This measure can be interpreted as the average magnitude of acceleration of landmarks while they traverse their estimated trajectories across time. A smaller MIR value would generally indicate smoother predicted landmark paths and is thus more biologically plausible because the lung would tend to move in a smooth fashion during ventilation.

Finally, to visually and qualitatively examine the result of registration, we generated temporal mean and variance of the final deformed moving tissue-volume image before and after registration. We expect accurate registration would sharpen the mean image and reduce the value of the variance image especially in regions where the mean image was originally blurry [1] [17]. 4D Jacobian image was also computed from the estimated 4D transform that deformed each phase within the original 4D tissue-volume image to match the extreme exhale phase. Temporal mean and variance of the 4D Jacobian image were overlaid on top of the extreme

exhale phase image to qualitatively illustrate lung ventilation behavior.

3. Experiments and Results

To assess the performance of the proposed algorithm, we performed experiments on 2D+t synthetic image and 4DCT pulmonary data sets, and compared the results of this algorithm with those of existing pairwise 3D SSTVD[18][5] and 4D SSD[10] methods. There exists a trade-off between registration accuracy and temporal smoothness, so we evaluated both of these factors to demonstrate that the proposed 4D SSTVD method achieves a good balance between them.

3.1. 2D+t Synthetic Data Experiment

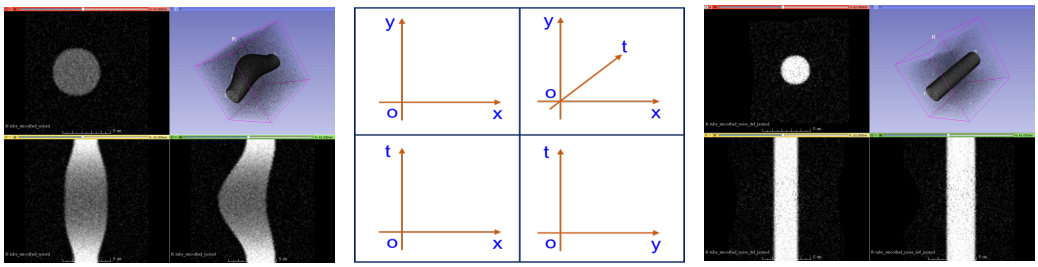
We tested the proposed 4D SSTVD algorithm on a 129x129x129-pixel 2D+t synthetic image as shown in Figure 2. The image consists of a time series of disk-shaped regions, whose centers' X coordinates remained the same across time while Y coordinates traced a periodic sinusoid trajectory in time. Conceptually, we can think of the 2D images within the time series as having already been converted into tissue-volume images and their intensity values are within range $[0, 1]$, representing fraction of tissue within voxels. The radii and intensities of the disks vary with respect to time but the total "amount of tissue" remained constant. This means the intensity is inversely proportional to the area of the disk at any time. For example, if the smallest disk at $t = 0$ has intensity I_1 and area A_1, then a disk later in the time series with area A_n will have intensity $I_n = A_1/A_n \cdot I_1$. The synthetic image was blurred by a Gaussian filter with standard deviation of 1 and Guassian noise was applied with 0 mean and 0.005 variance. Each sub-figure in Figure 2 consists of four views. The upper-left view is the disk at the middle time point. The lower-left and lower-right views are cross sections of the time series of disks parallel to x-t and y-t plane, respectively. The upper-right view is a 3D (2D+t) rendering of the synthetic image. The orientation of the four views of the synthetic image is as shown in Figure 2b.

This 2D+t synthetic image is intended to simulate a real 4DCT data set in the sense that the fraction of tissue within voxels varies periodically throughout the respiratory cycle but the total amount of tissue is conserved. Using the proposed 4D SSTVD method and the existing 4D SSD method, we performed registration on this 2D+t synthetic image and chose the 2D image at $t = 0$, whose disk-shaped region has the highest intensity and the smallest radius, as the target image. We used the same optimization parameters for both 4D SSTVD and 4D SSD methods, with a total of 8 resolutions in the multi-grid multi-resolution registration framework. Let's denote the 2D+t synthetic tissue-volume image as I, and the estimated transforms using 4D SSTVD and 4D SSD methods as $\mathbf{T}^r_{SSTVD}$ and $\mathbf{T}^r_{SSD}$, where $r \in \{0, 1, ..., 7\}$ is the resolution index. Then the deformed 2D+t synthetic tissue-volume image (without Jacobian modification) are $I(\mathbf{T}^r_{SSTVD})$ and $I(\mathbf{T}^r_{SSD})$. Through visual observation of Sub-Figure 2d-2f, we see that after resolution $r = 4$, the 4D SSTVD method already roughly aligned the time series and at finer resolutions, the results seemed to have stabilized. In Sub-Figure 2c, we applied proper Jacobian modification $|J_{\mathbf{T}^7_{SSTVD}}|$ onto the deformed moving image $I(\mathbf{T}^7_{SSTVD})$ to obtain the appropriate 2D+t deformed moving tissue-volume image $|J_{\mathbf{T}^7_{SSTVD}}| \cdot I(\mathbf{T}^7_{SSTVD})$, in which the 2D deformed moving tissue-volume images at other time points carry similar radius and intensity value to the target image at $t = 0$. In comparison, after resolution $r = 4$, the 4D SSD method didn't quite align the time series, as shown in Sub-Figure 2d. And when the 4D SSD registration proceeded to finer resolutions, the registration process seemed to have collapsed, resulting in biologically infeasible transformations, as shown in Sub-Figure 2h, 2i.

3.2. 4DCT Pulmonary Data Set Experiment

Quantitative experiments were performed on respiratory-gated 4DCT pulmonary data sets. Specifically, we used the publicly available POPI data set containing 4D landmarks (100 landmarks on each phase) for 3 patients, provided by J. Vandemeulebroucke et al.[15]. For each patient, the 4DCT image data consisted of 10 images corresponding to 10 breathing phases within the respiratory cycle. We compared the performance of pairwise 3D SSTVD, 4D SSD and the proposed 4D SSTVD tissue preservation algorithm. The pairwise 3D SSTVD registration scheme uses the extreme exhale phase image as an explicit target image, while the 4D SSD and proposed 4D SSTVD tissue preservation methods both use an implicit target image that is the temporal average of the 4D deformed moving tissue-volume image, which will be iteratively updated until convergence. No masks were used during registration. For all 3 POPI data sets, the B-spline grid spacings for 4D SSTVD and 4D SSD algorithms in three spatial dimensions are 96, 64, 64, 32, 32, 16, 16, 8 (mm) for forward transform and 64, 32, 32, 16, 16, 8, 8, 4 (mm) for inverse transform, corresponding to down-sampling ratios of the 4D image 1/8, 1/8, 1/4, 1/4, 1/2, 1/2, 1, 1 in spatial dimensions. The grid spacing is always 1 and no down-sampling is used in the temporal dimension. For 3D SSTVD algorithm, the B-spline grid schedule is 48, 32, 32, 16, 16, 8, 8, 4 (mm), with the same image down-sampling ratios in spatial dimensions as in the 4D algorithms.

Mean landmark error and mean irregularity, as introduced in previous subsection, were computed and the results are shown in Tables 1 and 2. The last row of Table 1 consists of the results obtained by Vandemeulebroucke el al.[15]. We can see that the proposed 4D SSTVD algorithm



(a) I (b) Orientation of four views (c) $|J_{\mathbf{T}^7_{SSTVD}}| \cdot I(\mathbf{T}^7_{SSTVD})$

(d) $I(\mathbf{T}^4_{SSTVD})$ (e) $I(\mathbf{T}^5_{SSTVD})$ (f) $I(\mathbf{T}^7_{SSTVD})$

(g) $I(\mathbf{T}^4_{SSD})$ (h) $I(\mathbf{T}^5_{SSD})$ (i) $I(\mathbf{T}^7_{SSD})$

Figure 2: Comparison between results from 4D SSTVD and 4D SSD algorithms to register 2D+t synthetic tissue-volume image I.

achieved better accuracy than 4D SSD method. At the same time, it achieved better temporal smoothness compared to 3D pair-wise SSTVD, which is measured by the average magnitude of acceleration of landmarks when they move along their estimated paths through the respiratory cycle. It should be noted that the relatively big difference among irregularity values for different data sets might stem from the possibility that landmarks for different data sets reside in different regions of the lung. For example, there may be more landmarks located around the diaphragm or other high functioning regions of the lung in POPI Patient 2 than in POPI Patient 1, causing the landmark irregularity to be generally higher for the former.

Visual inspection of the registration result is illustrated in Figure 3. Before registration, the temporal mean of the 4D tissue-volume image was blurry, especially near vessels and the diaphragm. And the temporal variance image had high values in these blurry areas. After registration using the proposed 4D SSTVD algorithm, the mean image became a lot sharper and the variance image grew a lot darker within the lung region, indicating that the registration indeed made the phase images better aligned. The bright stripe near the bottom in the variance image after registration result from the fact that the bottom regions of different phase images were moved up by different amounts to match the extreme exhale phase.

Accuracy (mm)	POPI Patient 1	POPI Patient 2	POPI Patient 3
Before Registration	3.44 ± 3.06	6.41 ± 6.09	3.65 ± 3.89
Proposed 4D SSTVD	0.83 ± 0.63	1.11 ± 0.89	0.87 ± 0.77
4D SSD	0.83 ± 0.65	1.43 ± 1.83	1.02 ± 1.09
3D SSTVD Pairwise	0.80 ± 0.63	1.21 ± 1.29	0.92 ± 0.89
Vandemeulebroucke et al.	0.96 ± 0.66	1.20 ± 0.96	1.11 ± 1.14

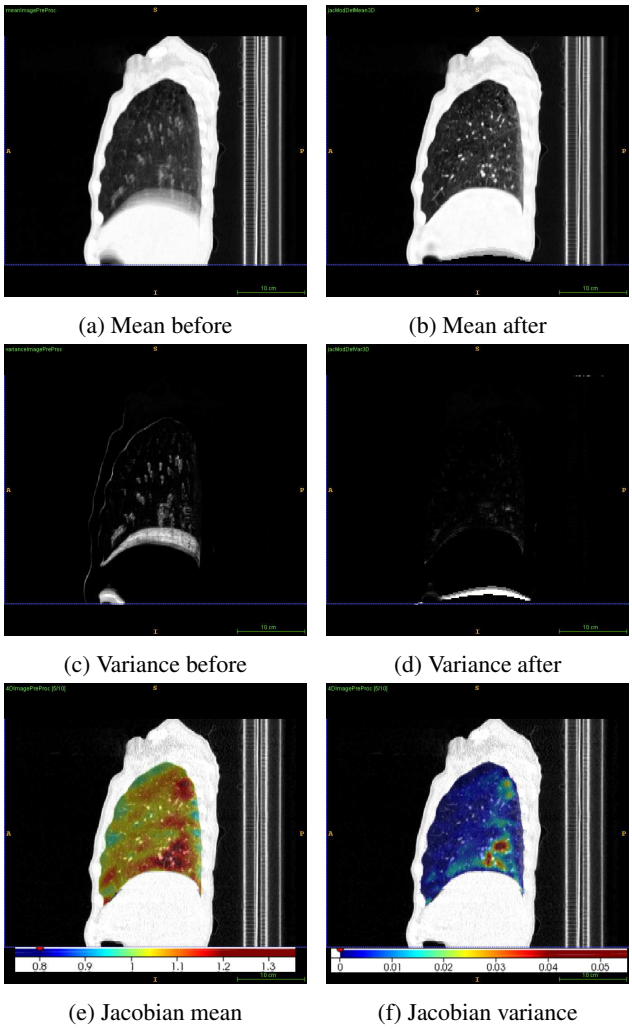
Table 1: Registration accuracy measured by average landmark error (mean $\pm$ standard deviation, smaller is better).

4D Jacobian image was also computed from the composed 4D transform deforming all phase images to match

Irregularity ($mm/phase^2$)	POPI Patient 1	POPI Patient 2	POPI Patient 3
Proposed 4D SSTVD	0.94 ± 0.57	2.38 ± 2.15	1.52 ± 1.25
4D SSD	0.94 ± 0.57	2.26 ± 1.98	1.45 ± 1.23
3D SSTVD Pairwise	1.04 ± 0.71	2.67 ± 2.26	1.71 ± 1.34

Table 2: Temporal smoothness measured by landmark trajectory irregularity (mean $\pm$ standard deviation, smaller is better).

the extreme exhale phase. 4D masks for the 4DCT data sets were generated using a 4D optimal surface finding (OSF) algorithm proposed by Gerard et al. [16]. The masks were applied onto the Jacobian images to mask out regions outside the lung, and the Jacobian values within the lung region would provide relevant information about pulmonary ventilation behavior. For our experiments on clinical 4DCT data sets using the proposed 4D SSTVD method, all Jacobian values obtained from the estimated 4D transform were positive, indicating no folding or collapsing of space was introduced by the transform and thus the predicted lung motion would be biologically feasible. In Sub-Figure 3e, the temporal mean of 4D Jacobian image was overlaid on the extreme exhale phase to qualitatively illustrate the average ventilation behavior of the lung starting from the extreme exhale phase. In Sub-Figure 3f, the temporal variance of 4D Jacobian image was overlaid on the extreme exhale phase. Larger values of Jacobian variance indicate bigger differences among the predicted expansions from the extreme exhale phase to each of the other phases. High Jacobian variance regions roughly correspond to the most blurry regions in Sub-Figure 3a.



(a) Mean before (b) Mean after

(c) Variance before (d) Variance after

(e) Jacobian mean (f) Jacobian variance

Figure 3: Comparison of temporal mean and variance of 4D deformed moving tissue-volume image before and after registration. temporal mean and variance of 4D Jacobian image overlaid on extreme exhale phase.

4. Conclusion and Discussion

The 4D tissue preserving algorithm inherits the advantage of 3D SSTVD to handle registration scenarios where spatially corresponding voxels have varying CT numbers due to changes in fraction of tissue within voxels while the total tissue volume is preserved. Meanwhile, the 4D cubic B-spline transformation model and temporally extended linear elasticity ensure the temporal smoothness of the deformation field. Comparison results on 4DCT data sets indicate the proposed 4D SSTVD algorithm strikes a good balance between accuracy and temporal regularity. Without an explicit target image in the 4D registration framework, all information of the dynamic data set is considered simultaneously, avoiding bias toward any specific reference image and increasing robustness against potential outliers caused by artifacts or noise[17]. By incorporating temporal information within 4DCT data sets, the proposed method can provide more relevant information for motion tracking and ventilation estimation and thus aid in radiotherapy treatment planning [4],[17]. Another benefit of considering temporal information is that temporal interpolation can be used to provide an estimate of lung motion at certain intermediate time points that are not present in the initial 4DCT data set.

5. Acknowledgment

This work is supported in part by National Cancer Institute of the National Institute of Health (NIH) under award numbers R01CA166703 and R01CA166119. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. We would like to thank Professor David Stewart from Mathematics department of University of Iowa for answering my questions about numerical optimization. We also want to thank lab mate Ali Ghayoor for answering my questions about ITK.

References

[1] S. K. Balci, P. Golland, M. Shenton, and W. M. Wells. Free-form b-spline deformation model for groupwise registration. In *Medical image computing and computer-assisted intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 10, page 23. NIH Public Access, 2007.

[2] G. E. Christensen and H. J. Johnson. Consistent image registration. *IEEE Trans. Med. Imaging*, 20(7):568–582, July 2001.

[3] G. E. Christensen, J. H. Song, W. Lu, I. El Naqa, and D. A. Low. Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. *Medical Physics*, 34(6):2155–2163, 2007.

[4] S. Flampouri, S. B. Jiang, G. C. Sharp, J. Wolfgang, A. A. Patel, and N. C. Choi. Estimation of the delivered patient dose in lung IMRT treatment based on deformable registration of 4d-ct data and monte carlo simulations. *Physics in Medicine and Biology*, 51(11):2763, 2006.

[5] V. Gorbunova, J. Sporring, P. Lo, M. Loeve, H. A. Tiddens, M. Nielsen, A. Dirksen, and M. de Bruijne. Mass preserving image registration for lung ct. *Medical Image Analysis*, 16(4):786–795, 2012.

[6] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim. Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.

[7] G. Li, D. Citrin, K. Camphausen, B. Mueller, C. Burman, B. Mychalczak, R. W. Miller, and Y. Song. Advances in 4d medical imaging and 4d radiation therapy. *Technology in cancer research & treatment*, 7(1):67–81, 2008.

[8] D. C. Liu and J. Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[9] J. R. McClelland, J. M. Blackall, S. Tarte, A. C. Chandler, S. Hughes, S. Ahmad, D. B. Landau, and D. J. Hawkes. A continuous 4d motion model from multiple respiratory cycles for use in lung radiotherapy. *Medical Physics*, 33(9):3348–3358, 2006.

[10] C. Metz, S. Klein, M. Schaap, T. van Walsum, and W. J. Niessen. Nonrigid registration of dynamic medical imaging data using nd+t b-splines and a groupwise optimization approach. *Medical Image Analysis*, 15(2):238–249, 2011.

[11] J. M. Reinhardt, K. Ding, K. Cao, G. E. Christensen, E. A. Hoffman, and S. V. Bodas. Registration-based estimates of local lung tissue expansion compared to xenon-CT measures of specific ventilation. *Medical Image Analysis*, 12(6):752–763, 2008.

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

[13] M. Schaap, C. T. Metz, T. van Walsum, A. G. van der Giessen, A. C. Weustink, N. R. Mollet, C. Bauer, H. Bogunović, C. Castro, X. Deng, et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical Image Analysis*, 13(5):701–714, 2009.

[14] H. Sundar, H. Litt, and D. Shen. Estimating myocardial motion by 4d image warping. *Pattern Recognition*, 42(11):2514–2526, 2009.

[15] J. Vandemeulebroucke, S. Rit, J. Kybic, P. Clarysse, and D. Sarrut. Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. *Medical Physics*, 38(1):166–178, 2011.

[16] S. G. Yeary, G. E. Christensen, J. E. Bayouth, S. Bodduluri, Y. Pan, J. Guo, K. Du, J. H. Song, B. Zhao, I. Oguz, and J. M. Reinhardt. 4D lung CT segmentation for radiation therapy applications. In *ICART: Imaging and Computer Assistance in Radiation Therapy*, pages 50–57, 2015.

[17] M. Yigitsoy, C. Wachinger, and N. Navab. Temporal groupwise registration for motion modeling. In *Information Processing in Medical Imaging*, pages 648–659. Springer, 2011.

[18] Y. Yin, E. A. Hoffman, and C.-L. Lin. Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Medical Physics*, 36(9):4213–4222, 2009.