Deformable Image Registration for Breath-hold CT Image Pairs from the COPDgene Study

Edward Castillo Jr. Ph.D.\textsuperscript{1,4}, Richard Castillo M.Sc.\textsuperscript{2}, Xuejun Gu Ph.D.\textsuperscript{5}, Josue Martinez Ph.D.\textsuperscript{6}, David Fuentes Ph.D.\textsuperscript{2}, Steve Jiang Ph.D.\textsuperscript{5}, Paul Friedman M.D.\textsuperscript{5}, and Thomas Guerrero M.D. Ph.D.\textsuperscript{1,3,4}

Departments of \textsuperscript{1}Radiation Oncology and \textsuperscript{2}Imaging Physics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030; \textsuperscript{3}Graduate School of Biomedical Sciences, The University of Texas, Houston, Texas, 77225; \textsuperscript{4}Department of Computational & Applied Mathematics, Rice University, Houston, Texas 77251; \textsuperscript{5}Department of Radiology, University of California San Diego, La Jolla, California 92093; \textsuperscript{6}Department of Epidemiology & Biostatistics, Texas A & M University, College Station, Texas 77843;

Reprint requests to: Thomas Guerrero, M.D., Ph.D., Department of Radiation Oncology – Unit 97, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston TX 77030, e-mail address: tguerrero@mdanderson.org, telephone: 713-563-2300, fax: 713-563-2366.

Running title: DIR for COPDgene

submitted to: MEDICAL IMAGE ANALYSIS
Conflict of Interest Notification

The authors have no commercial or financial interests related to this study to disclose.
Abstract

The CODPgene study acquires maximum effort inhale and exhale breath-hold computed tomography (BH-CT) image pairs, with the exhale image acquired at one-fourth dose to reduce the radiation dose. The poor statistical quality of the exhale images and the extreme anatomic deformations due to the maximum inhalation make deformable image registration (DIR) of COPDgene images difficult. In this study, we introduce a new DIR algorithm, the Moving Least Squares Guided Local Optimization (MILO) method, designed to register noisy image pairs depicting drastic deformations. Rather than search for mass conservation or voxel similarity, MILO is formulated on the assumption that there is a change in intensity between corresponding voxels in the reference and target image that is unknown. The resulting equations are solved without explicitly solving for the unknown intensity differences by a local optimization approach. Using 5 COPDgene BH-CT image pairs with 4378 landmark point pairs (available at www.dir-lab.com), we compare the spatial accuracy performance of MILO versus the well known demons algorithm. Over the 5 COPDgene test cases MILO achieved an average spatial accuracy of 1.27 mm versus 21.37 mm for demons. To our knowledge, no other DIR algorithm is capable of accurately registering the BH-CT image pairs contained in the COPDgene data set.

Keywords: deformable image registration, computed tomography, chronic obstructive pulmonary disease, local optimization.
1. Introduction

Deformable image registration (DIR) is a cross-cutting technology with diagnostic and therapeutic medical applications. DIR algorithms were first developed in computer vision research to estimate motion by warping a source image onto a target, producing an estimated image that visually appeared similar to the target image (Horn and Schunck, 1981). Though their spatial accuracy was unknown, DIR algorithms were applied to extract physiological information such as cardiac wall motion (Song and Leahy, 1991) and ventilation (Guerrero et al., 2005) from medical images. For medical applications the goal in applying DIR is to obtain an accurate spatial registration of the underlying anatomy and not simply to achieve image similarity. We developed a statistical framework to assess DIR spatial accuracy using large sets of manual landmark points (Castillo et al., 2009b) and made our study data sets publicly available (www.dir-lab.com). The optimal algorithm and its spatial accuracy in registering the underlying anatomy should be assessed for each specific application.

Chronic obstructive pulmonary disease (COPD) is the fourth leading cause of death in the United States and is projected to be the third leading cause of death worldwide by the year 2020 (Lopez and Murray, 1998). The COPDgene study (www.copdgene.org) is a National Heart Lung Blood Institute (NHLBI) funded cross-sectional study, with a recruitment goal of over 10,000 patient, designed to discover what genetic factors contribute to the development of COPD. Disease severity is documented using pulmonary function testing and inspiratory & expiratory breath-hold computed tomography (i & eBH-CT) imaging. The iBH-CT is acquired at maximum effort and the eBH-CT at end normal expiration, called the functional residual capacity (FRC). The eBH-CT images are acquired at one-fourth the CT tube current as the iBH-CT in order to
reduce the radiation dose to the study subjects, as a consequence the eBH-CT image noise is doubled. Reports on the evaluation of i- & eBH-CT image pairs for COPD are limited to the evaluation of each of the eBH- & iBH-CT images separately (Akira et al., 2009; Kubo et al., 1999; Yamashiro et al., 2010; Zaporozhan et al., 2005) as there is no validated algorithm which can accurately register these image pairs for their joint analysis.

DIR registration of the COPDgene BH-CT image pairs is challenging due to the large displacements, change in density and CT value, the difference in image noise, the highly non-uniform mechanical properties of lung tissue with COPD, and the changes in anatomic shape of the vasculature due to the large volume change. These pronounced differences are illustrated in Figures 1 and 2. There are currently no validated DIR tools available to provide a link between the COPDgene BH-CT image pairs. New algorithms are required to achieve high spatial accuracy and avoid convergence on local minima for these image pairs. In this study, we create a reference data set consisting of 5 pairs of BH-CT image pairs from the COPDgene study. We identify and statistically characterized sets of landmark point pairs for each image pair to use in the evaluation of DIR algorithms. This data is publically available at www.dir-lab.com so other investigators can compare alternative algorithms. We introduce a new DIR algorithm, Moving Least Squares Guided Local Optimization (MILO), designed to register noisy image pairs depicting drastic deformations. In this study, MILO is compared with the well known demons algorithm (Gu et al., 2010; Thirion, 1998) for spatial accuracy.

This paper is organized as follows: Section 2 briefly describes the MILO and demons algorithm. The methods used in comparing the spatial accuracy between the two DIR algorithms is described next. The image data and expert-determined landmark validation data set are
presented. Section 3 provides the mathematical theory and computational methods used in the MILO algorithm. Section 4 presents the results of the spatial accuracy assessment and comparison. Section 5 is the discussion and section 6 the conclusion.

2. Materials and Methods

2.1.1 Demons DIR algorithms

The demons DIR algorithm uses local forces applied by ‘demons’ that displace the voxels in the target image to match the reference image. The target image is iteratively deformed by applying a displacement vector in conjunction with a regularization term. The displacement vector for each voxel is estimated by inexactly solving an incompressible optical flow equation. The Gaussian (or smoothing) regularization term is chosen to compensate for the inexact solution. The calculation of displacement vector of each voxel and a local smoothing filter are highly data-parallel computational tasks and well-suited for data-structure parallelization. With the aid of advanced CUDA environment on GPU, an ultrafast GPU-based parallel demons algorithm and its variations have been developed and accessed by Gu et al. (Gu et al., 2010). In this study, we utilize the passive-force demons DIR, which achieved the highest spatial accuracy among 6 demons implementations studied in our prior study (Gu et al., 2010). Details about the implementation and evaluation of demons algorithms can be found in the cite reference by Gu et al. (Gu et al., 2010).

2.1.2 MILO Method for DIR
In formulating the MILO DIR, we make the general assumption that the change in intensity between corresponding voxels in the reference and target image is additive and unknown. The full MILO algorithm is then based on iteratively solving a sequence of small, local, nonlinear least squares problems defined on a subset of all voxels contained in the image domain with the initial guess for each local subproblem provided by a Moving Least Squares (MLS) estimate. This local solution minimizes local mismatch. The final globally dense displacement field DIR solution is then the MLS interpolant of the local optimal solutions. The mathematical derivation and computational implementation of the MILO algorithm is given in section 3 below.

2.2 Spatial accuracy assessment

2.2.1 4DCT thoracic images

Five patients, treated for thoracic malignancies in the Department of Radiation Oncology at The University of Texas M. D. Anderson Cancer Center, and whose treatment planning 4DCT images have been previously utilized as part of a series of DIR accuracy evaluation studies were selected (Castillo et al., 2010a; Castillo et al., 2009a; Castillo et al., 2009b; Gu et al., 2010). Each patient had undergone a treatment planning session in which a 4DCT image of the entire thorax and upper abdomen was obtained at 2.5 mm slice spacing with a General Electric Discovery ST PET/CT scanner (GE Medical Systems, Waukesha, WI). The images were acquired with the patients in the supine position with normal resting breathing. The extreme inhalation and exhalation component phase images were utilized in this study. As previously reported, each image was cropped to include the entire ribcage and subsampled in the transverse plane (Castillo
et al., 2009b). The image characteristics of the five 4DCT cases utilized in this study are given in Table 1.

2.2.2 COPDgene CT images

Five maximum effort BH-CT image pairs were selected from COPDgene study cases. Each patient had received CT imaging of the entire thorax in the supine position at normal expiration and maximum effort full inspiration on the COPDgene study. The CT imaging was performed with a GE VCT 64-slice scanner (GE Healthcare Technologies, Waukesha, WI) with a pitch of 1.375 mm, speed of 13.75 mm per rotation, 120 kVp, 0.5 sec per rotation, 400 mA per rotation for inhale BH, and 100 mA per rotation for exhale BH. The images used in this study were reconstructed with a high resolution (BONE), with the lung diameter setting the field of view, and with 2.5-mm slice spacing. The image characteristics of the 5 COPDgene cases utilized in this study are given in Table 1.

2.2.3 Landmark selection and characterization

Measurements of DIR spatial accuracy was performed using manually identified sets of prominent anatomical landmark feature pairs identified across the maximum inhalation and exhalation BH-CT image pairs. A Matlab-based software interface named APRIL (Assisted Point Registration of Internal Landmarks), previously described (Castillo et al., 2009b), was utilized to facilitate manual selection of landmark feature pairs between volumetric images. Basic features of the software include separate window and level settings for each display, visualization of equivalent voxel locations in the orthogonal plains, and interactive tools for segmentation of lung voxels from the image data. To determine corresponding feature points the user must manually designate the feature correspondence via mouse click on the target image. No implanted fiducials
or added contrast agents were used to aid in the selection of landmark features, which typically included vessel and bronchial bifurcations. The manual selection and statistical characterization of the 4DCT reference data utilized in this study has been described previously in detail (Castillo et al., 2009b).

For the COPDgene datasets, source feature points were selected systematically on the 5 test image pairs by an expert in thoracic imaging, beginning at the apex of the lung. Points were selected with an initial goal of >5 feature points for each lung per axial image slice. This approach ensured the collection of >600 validation point pairs for each case distributed throughout the lungs. Following feature selection for a given case, all landmark pairs were visually reviewed by the primary reader a second time and the locations adjusted on the exhale image if necessary. The verification step was a required part of the initial registration process performed by the primary reader. A subset of 150 points were re-registered by the primary reader, to estimate intra-observer variance, and by secondary readers, to estimate inter-observer variance. For each of the 150 points given on the inhale BH-CT the re-registration process consisted of finding the corresponding point on the exhale BH-CT. Access to this dataset is available through the Internet (www.dir-lab.com). The full sets of points were then used to assess the spatial accuracy of DIR algorithms for this study. Characteristics of all reference datasets utilized in this study are given in Table 1.

2.3 Statistical methods

The primary endpoint of this study is the spatial accuracy assessment of the MILO DIR formulation applied to the COPDgene breath-hold image pairs. As a point of reference, we have included a passive force demons algorithm which has been previously shown to yield highly
accurate DIR for 4DCT treatment planning images (Gu et al., 2010). Comparative evaluation of MILO and demons DIR is performed separately for the 4DCT and COPDgene datasets. For both subsets of patient images, the nonparametric Wilcoxon rank sum test was used to compare the three-dimensional spatial accuracy results between methods. Additionally, the subset of landmark features sampled for repeat registration by the multiple expert observers was matched between all observers and both algorithms, in order to directly test the statistical significance of the difference in measured registration errors between each algorithm and the combined set of human observers. The Wilcoxon rank sum test was similarly applied for this purpose.

Summary statistics are provided for individual right-left (RL), anterior-posterior (AP), and superior-inferior (SI) component displacements, as well as the 3D-Euclidean displacements. The Wilcoxon rank-sum test was used to compare the mean spatial errors between the DIR algorithms for each, and across all five cases. All tests were two-sided, with \( p \leq 0.05 \) considered significant. Statistical tests which resulted in \( p \)-values less than \( 1 \times 10^{-6} \) were reported as \( p < 1 \times 10^{-6} \). Statistical analysis was done using statistical Analysis Systems, version 9 (SAS Institute, Cary, NC), and S-Plus, version 7 (Insightful, Seattle, WA).

3. Theory and Computational Methods

3.1 MILO: Overview - MLS Guided Local Optimization (MILO)

As stated earlier, the two primary DIR difficulties associated with the COPDGene data set are the poor statistical quality of the exhale images and the extreme deformations depicted in the image pairs. Image noise and large voxel displacements are known limitations for DIR methods (Barron et al., 1994). In addition, there is also significant changes in the shape of corresponding
anatomical structures (as shown in figure 2) as result of either the large displacements or the presence of COPD disease. The mass conservation assumption, recently reported in formulating DIR algorithms for 4DCT thoracic image sets (Castillo et al., 2010a; Castillo et al., 2009a; Yin et al., 2009), is also inappropriate due to a significant mass differences brought on by pulmonary blood volume changes between the exhale and inhale BH images. There is no a priori relationship between the intensities in the reference image and target image due to the spatially varying changes in both ventilation and blood content within the lungs.

The DIR method introduced here, referred to as the MILO DIR algorithm, is a computationally inexpensive and parallelizable algorithm designed to address the difficulties associated with the COPDGene dataset. The MILO method is an iterative procedure based on a local-global strategy that utilizes Moving Least Squares (MLS) to define a globally dense displacement field from the solutions to a series of localized DIR (nonlinear least squares) subproblems. In addition to the obvious benefit of being a parallelizable approach, decoupling the DIR problem into a series of subproblems allows us to model local intensity variations between the reference and target images as unknown quantities, without introducing auxiliary variables into the DIR problem. However, considering the potential for extreme deformations and the nonlinear nature of the optimization subproblems, the accuracy of the subproblem solutions are dependent on the quality of the initial guess provided to the optimization routine. Globally, the DIR solution is assumed to obey a simple ansatz: The forward (reference to target) deformation map is consistent with the backward (target to reference) deformation map. The MILO method iteratively computes both maps simultaneously, using the forward and backward
MLS estimates from the previous iteration to provide the initial guess to the DIR subproblems for the current iteration. The iteration ceases when the consistency ansatz is satisfied.

3.2 Localized Nonlinear Least Squares DIR Subproblem: General Intensity Variation Model

As opposed to solving a single large-scale optimization problem, the MILO method is based on decoupling the full DIR problem into a series of subproblems. Each subproblem utilizes local image information to describe the displacement of a single voxel as the least squares fit to a general intensity variation model.

Let $R$ and $T$ be the scalar valued reference and target image respectively, and $\phi(x) \in \mathbb{R}^3$ be the unknown displacement field. In order to account for non-uniform intensity differences between the reference and target images we make the general assumption that the intensities variations are additive and unknown. Specifically, for the voxel $x$ in the reference image we assume:

$$R(x) = T(x + \phi(x)) + \Lambda(x),$$

(1)

where $\Lambda$ is an unknown scalar function in addition to the unknown displacement field $\phi$. At first glance, this formulation appears to be computationally burdensome since it would require the additional cost of solving for the intensity variation $\Lambda(x)$, whereas most DIR methods assume voxel intensities do not change from reference to target ($\Lambda \equiv 0$) and need only solve for $\phi$. However, model (1) can be implemented without explicitly solving for $\Lambda$ by adopting a local optimization approach. Consider the voxel $x_i$ in the reference image and let $\Omega_i(x_i)$ be a
neighborhood centered at $x_i$ with radius $\varepsilon$. The optimal (in the least squares sense) displacement vector $d_i$ and intensity variation $\lambda_i$ over the neighborhood $\Omega_i(x_i)$ are given by:

$$
(d_i, \lambda_i) = \arg \min_{(d, \lambda)} G(d, \lambda),
$$

$$
G(d, \lambda) = \frac{1}{2} \sum_{x_j \in \Omega_i(x_i)} \left( T(x_j + d) - R(x_j) + \lambda \right)^2.
$$

(2) is a straightforward nonlinear least squares problem but considering the poor quality of the target image and the effects of drastic deformations, solving problem (2) requires a robust optimization algorithm. Typical gradient based methods such as Gauss-Newton’s Method or the Levenberg-Marquardt method require a continuous representation of the image and are sensitive to the image noise (Bjorck, 1996).

The fact that image data is inherently defined on a grid suggests the employment of a grid based search algorithm for solving problem (2). The search algorithm approach is simple; merely evaluate the objective function at each grid point within a specified target window and choose the location yielding the smallest value. Though this approach is well suited for the displacement variable $d$, the intensity variation variable $\lambda$ is independent of the image grid discretization. However, examination of the objective function $G$ reveals that for any fixed value of $d$, the value of $\lambda$ yielding the smallest objective value is given by the mean intensity mismatch:

$$
\lambda = \frac{1}{|\Omega_i(x_i)|} \sum_{x_k \in \Omega_i(x_i)} T(x_k + d) - R(x_k),
$$

where $|\Omega_i(x_i)|$ is the number of voxels contained in the $\varepsilon$ neighborhood of $x_i$. As a corollary, the first order optimality condition for problem (2) indicates that at a critical point, definition (3) must hold. This result allows for the definition of a merit function
\[
\hat{G}(d) = \frac{1}{2} \sum_{j \in \Omega(x_j)} \left( T(x_j + d) - R(x_j) - \left[ \frac{1}{\Omega(x_j)} \sum_{k \in \Omega(x_k)} T(x_k + d) - R(x_k) \right] \right)^2, \quad (4)
\]

that when used in place of \( G \) within the grid search algorithm removes the need to explicitly solve for \( \lambda \). As a result, utilizing (4) within the grid search algorithm constitutes a simple optimization solver for problem (2) and allows for the utilization of the general intensity model (1) without augmenting the number of unknown in the DIR problem. Naturally, formulation (2) is easily adjusted for mapping target image voxels into the reference image:

\[
(d_i, \lambda_i) = \arg \min_{(d, \lambda)} H(d, \lambda),
\]

\[
H(d, \lambda) = \frac{1}{2} \sum_{j \in \Omega(x_j)} \left( R(x_j + d) - T(x_j + \lambda) \right)^2. \quad (5)
\]

As would also be the case for gradient methods, it is possible for the search method to find solutions corresponding to erroneous local minima. Though this is a common issue for nonlinear, non-convex optimization, further compounding the problem are the effects of extreme deformations (see figure 2) and image noise, which can cause even the global minimizer for a subproblem to be physically inaccurate (two unrelated target and reference image locations can look alike). Robustness to erroneous local minima requires that a high quality initial guess be provided to the optimization solver.

3.3 Globally Defined Moving Least Squares Deformations Function From Local Solutions

Moving Least Squares (MLS) is a highly generic and versatile tool for approximating an unknown function by fitting polynomials to function samples given at uniform or non-uniform
locations (Bos and Salkauskas, 1989; Lancaster and Salkauskas, 1981). Though most commonly employed to reconstruct surfaces from noisy, unstructured point cloud data, MLS has recently found utility in DIR (Castillo et al., 2010a; Schaefer et al., 2006). Within the MILO framework, MLS is used to define a globally dense displacement field from a given set of voxel locations and their corresponding solutions to problem (2). Specifically, consider the voxel set $X = \{x_i\}_{i=1}^{N}$ and a corresponding set of displacement vector $D = \{d_i\}_{i=1}^{N}$. The MLS defined deformation function for $X, D$ is given by:

$$\phi_{\text{MLS}}(x; X, D) = p^*(x),$$

where the polynomial $p^*$ is the solution to the following linear least squares problem:

$$\min_{p \in \Pi_m} \sum_{i=1}^{N} \| p(x_i) - y_i \|^2 W(\| x_i - x \|),$$

where $W$ is a radial weighting function, and $y_i = x_i + d_i$.

The MLS minimization problem (7) is over the space $\Pi_m$ containing all polynomials of degree $m$, and the weighting function $W$ is typically chosen to be a smooth, fast decaying function. Here, $W(r) = e^{-\frac{r^2}{h^2}}$ with finite support of size $h$. Consequently, MLS allows for the description of a globally dense displacement field without requiring a large-scale linear system solve. Furthermore, the MLS field is locally defined as a least squares fit to the given input data and therefore inherently mitigates the effects of data noise, which in our case is caused by erroneous local minima associated with the DIR subproblems (2) and (5).

### 3.4 MLS Guided Local Optimization
While subproblems (2) and (5) model voxel motion at the local neighborhood level, we introduce a consistency ansatz to model motion at the global level. Specifically, consider a set of uniformly spaced voxels in the reference image \( X = \{ x_i \}_{i=1}^N \) and a corresponding set of displacement vectors \( D^F = \{ d_i^F \}_{i=1}^N \) that give the mapped position of each \( x_i \) within the target image as \( x_i + d_i^F \). The forward (reference to target) MLS deformation function defined by \( X, D^F \) is given by \( \phi_{mls}(x; X, D^F) \). Similarly, let \( Y = \{ y_j \}_{j=1}^M \) be a set of uniformly spaced target image voxels with corresponding displacements \( D^B = \{ d_j^B \}_{j=1}^M \) that give the mapped position of each \( y_j \) within the reference image as \( y_j + d_j^B \). The backward (target to reference) deformation function defined by \( Y, D^B \) is given by \( \phi_{mls}(y; Y, D^B) \). The consistency ansatz is then

\[
\begin{align*}
\phi_{mls}(x; X, D^F) &= \phi_{mls}(x; \hat{X}, -D^B), \forall x \in X, \\
\phi_{mls}(y; Y, D^B) &= \phi_{mls}(y; \hat{Y}, -D^F), \forall y \in Y,
\end{align*}
\]

where the set \( \hat{X} = \{ \hat{x}_j \}_{j=1}^M \) with \( \hat{x}_j = y_j + d_j^B \) and \( \hat{Y} = \{ \hat{y}_i \}_{i=1}^N \) with \( \hat{y}_i = x_i + d_i^F \). Put simply, the MLS function defined by the forward displacements should be consistent with the one defined by the backward displacements, an assumption commonly used within the context of registration of anatomical structure (Christensen and Johnson, 2001). Additionally, the ansatz serves as a type of merit functional for the MLS fields generated by the solutions to subproblems (2) and (5), and thus provides robustness to erroneous, spatially inaccurate local minima.

For the given voxel sets \( X \) and \( Y \), the MILO method computes displacement vector sets \( D^F \) and \( D^B \) satisfying the ansatz (8) such that for each \( x_i \in X \), the corresponding vector \( d_i^F \) is a local minimizer for the nonlinear least squares problem (2) defined by \( x_i \), and for each \( y_i \in Y \),
the corresponding vector \( d_i^B \) is a local minimizer for the nonlinear least squares problem (5) defined by \( y_i \). The MILO DIR solution is then given by either the forward or backward MLS mappings \( \phi_{mls}(x; X, D^F) \) and \( \phi_{mls}(y; Y, D^B) \). Algorithmically, the solution sets \( D^F, D^B \) are determined simultaneously in an iterative fashion. At each iteration, the estimated mappings from the previous iteration are used to determine the initial guesses supplied to the optimization solver for the subproblems defined by each \( x_i \in X \) and \( y_j \in Y \). The iteration ceases when ansatz (8) is satisfied.

3.5 The MILO Algorithm

Given the voxel set \( X, Y \), and the parameters \( \epsilon \) (neighborhood radius for DIR subproblems) and \( h \) (support of the MLS weighting function), define the set of forward displacement estimates \( \tilde{D}^F = \{ \tilde{d}_i^F \}_{i=1}^N \) for \( X \), and backward displacement estimates \( \tilde{D}^B = \{ \tilde{d}_i^B \}_{i=1}^M \) for \( Y \).

1. For each \( x_i \in X \), determine \( d_i^F \) by solving problem (2) with initial guess \( \tilde{d}_i^F \).

2. For each \( y_j \in Y \), determine \( d_j^B \) by solving problem (5) with initial guess \( \tilde{d}_j^B \).

3. If ansatz (8) holds, return \( D^F \) and \( D^B \) as the solution sets. Otherwise update \( \tilde{D}^F \) and \( \tilde{D}^B \) and go back to step 2.

The update procedure utilized in step 3 simply sets the new estimate for each voxel to the midpoint between given by the forward and backward displacement estimates:

\[
\tilde{d}_i^F = \frac{1}{2} \left( \phi_{mls}(x_i; X, D^F) + \phi_{mls}(x_i; \hat{X}, -D^B) \right),
\]

\[
\tilde{d}_j^B = \frac{1}{2} \left( \phi_{mls}(y_j; Y, D^B) + \phi_{mls}(y_j; \hat{Y}, -D^F) \right).
\]
3.6 Numerical Implementation

Our current parallel implementation of the MILO method was written in C++ using the Trilinos Scientific Software Collection (Heroux et al., 2003). The parameter \( h \) (the MLS weighting function support) is allowed to change dynamically, a common practice in surface reconstruction settings, according to k-nearest neighbors in terms of the millimeter spacing of the image pairs. The MLS polynomial space used for all experiments was \( \Pi_1 \) (affine functions).

Additionally, an initial location guess is performed to determine the displacement estimates \( \tilde{D}^F \) and \( \tilde{D}^B \) for the first iteration of the MILO method by examining the spatial density distribution of each image. A cumulative density distribution (CDD) is formed along each axis for the iBH- and eBH-CT images. Each CDD is a discrete representation of a monotonically increasing function. The estimated motion in one axial direction is then determined by solving a one-dimensional registration problem mapping the corresponding CDD functions.

4. Results

4.1 Patient Characteristics

The image and DIR reference landmark characteristics for all cases included in this study are summarized in Table 1. The data acquisition and statistical characterization of the DIR reference data for the thoracic 4DCT image sets has been previously reported in detail (Castillo et al., 2009b). Each COPDgene image set was reconstructed to image dimensions \( 512 \times 512 \times N \) (range: 102-131), with in-plane voxel dimensions ranging from \((0.590 \times 0.590) - (0.652 \times 0.652)\) mm, and fixed 2.5 mm slice thickness. For each COPDgene image set, a minimum of 618 (max:
1172) anatomical landmark features was identified and manually registered between the maximum effort BH-CT images. Repeat registration of sampled subsets of landmarks for each case yielded estimates of observer registration error, which ranged in average from 0.58 (SD: 0.87) to 1.06 (SD: 1.51) mm for each case. Average three-dimensional Euclidean landmark displacements varied substantially among cases, ranging from 12.29 (SD: 6.39) to 30.90 (14.05) mm. Figure 3 shows the set of reference displacement vector field projections for each case, in oblique (top) and lateral (bottom) aspects, overlaid the maximum effort BH-CT isosurface. The color scale shown at right is fixed for each projection, illustrating the variability in motion field characteristics among the five patient cases.

4.2.1 4DCT spatial accuracy assessment

Table 2 shows a complete summary of the measured DIR spatial errors for both MILO and demons algorithms, and all patient cases included in this study. Measured errors are shown in component RL, AP, and SI directions, as well as in three-dimensional Euclidean magnitude. The demons DIR spatial accuracies for the thoracic 4DCT images have been previously reported (Gu et al., 2010) and are included here for comparative evaluation with MILO DIR over the same patient set. For registration of the 5 thoracic 4DCT cases, DIR for both algorithms consistently resulted in improved spatial alignment of the reference landmark features. Mean (and standard deviation) Euclidean errors ranged from 1.04 (1.15) - 2.51 (2.49) mm for demons DIR, and 0.69 (0.96) - 1.26 (1.50) mm for MILO DIR. For the combined set of 4DCT data ($N = 6762$ reference measurements), mean spatial registration error for demons DIR was 1.53 (1.58) mm, and 0.92 (1.14) mm for MILO DIR. Due to the large skewness in the data, for statistical
analysis of the measured errors we first transform the error outcomes using a $\log(x+1)$ function. The form of the data prompted us to use the non-parametric Wilcoxon rank sum test to compare the results between methods. Results of the test are shown in Table 3, and indicate that the MILO DIR performed with significantly ($p$-value < 0.05) less spatial registration error than demons DIR for registration of the thoracic 4DCT images. Moreover, we see that the observer errors measured by repeat registration on the 4DCT images are statistically equivalent to that of the MILO DIR, ($p$-value > 0.05). These results reflect comparisons for each of the 5 subjects individually, and using all subjects simultaneously in the analysis. All results presented in Table 3 were corrected for multiple comparison using the procedure of Hochberg (Hochberg and Tamhane, 1987).

4.2.2 COPDgene spatial accuracy assessment

All measurements of DIR spatial accuracy obtained over the 5 COPDgene patient sets are summarized in Table 2. As with the 4DCT images, MILO DIR consistently resulted in improved alignment of the reference landmark features, with mean (and standard deviation) Euclidean magnitude errors ranging from 0.96 (1.10) to 2.47 (3.67) mm. In contrast, the demons DIR numerically led to increased mis-registration in one instance (COPD-2) out of five, with magnitude errors ranging in average from 7.22 (5.30) to 29.45 (13.01) mm. For the combined set of COPDgene data ($N = 4378$ reference measurements), mean spatial registration error for demons DIR was 21.37 (10.20) mm, and 1.27 (1.95) mm for MILO DIR. Results of the Wilcoxon rank sum test are summarized in Table 3. As with the 4DCT images, the MILO DIR resulted in significantly ($p$-value < 0.05) less spatial registration error for all cases compared to
demons DIR. However, contrary to the 4DCT cases, MILO DIR was unable to achieve COPDgene spatial errors that were statistically indistinguishable from the combined observer errors obtained by repeat registration, \((p\text{-value} \geq 0.17)\).

Figure 4 shows graphical and pictorial representations of the DIR error assessment for both algorithms over an example case included in this study (COPD-5). In Figure 4a, histograms are shown for both algorithms in which spatial registration errors have been binned into 5 mm increments for the 4378 COPDgene BH-CT reference measurements. The calculated MILO displacement vectors corresponding to the reference landmark positions are shown in Figure 4b, overlain the iBH-CT isosurface, while the corresponding spatial registration errors are shown in Figure 4c, overlain the eBH-CT isosurface. For each reference landmark, error vectors are shown pointing from the calculated position in the eBH-CT image to that determined manually by the expert observer. Similar plots are shown in Figures 4d & e for the demons DIR. It is clear from figure 4e that a large number of calculated demons displacements landed outside the anatomic target, yielding large spatial errors in the range \(\geq 40\) mm (color scale indicated).

Figure 5 shows a box plot summary of all spatial registration error measurements obtained in this study for the MILO and demons algorithms, grouped into 4DCT (Figure 5a) and COPDgene (Figure 5b) datasets. In both cases, the figure graphically depicts a considerable difference in the distribution of measured errors between methods, consistent with the statistical assessment summarized in Table 3.
5. Discussion

In this study we introduce a novel deformable image registration algorithm for breath-hold CT image pairs obtained as part of the National Heart Lung Blood Institute (NHLBI) sponsored COPDgene study (www.copdgene.org). Three major multi-site NHLBI-sponsored studies (Lung Tissue Research Consortium (LTRC), COPDgene, and Sub-Populations and InteRmediate Outcome Measures in COPD Study (SPIROMICS)) collect and archive exhale and inhale BH-CT image pairs from thousands of research study subjects along with clinical, tissue, and genomic data (Beek and Hoffman, 2009; Holmes III et al., 2006; Punturieri et al., 2008; Regan et al., 2010). These studies were each designed with an imaging archival core to make the study CT images available to other investigators who may develop novel image derived biomarkers to characterize COPD disease or its consequences. DIR registration of the COPDgene BH-CT image pairs is challenging due to the large change in density and CT values (Figure 1), the changes in anatomic shape of the vasculature due to the large volume change (Figure 2), the large displacements (Figure 3), the difference in image noise (Figure 1), and the highly non-uniform mechanical properties of lung tissue with COPD (Figure 6). The MILO algorithm was designed to address these challenges.

In order to account for the voxel intensity variations between the reference and target image, the MILO algorithm utilizes a general voxel intensity model that treats the intensity variation as an unknown additive quantity without requiring the introduction of auxiliary variables. Moreover, the MILO method is based on a unique local-global strategy, similar to the approaches described in (Bruhn et al., 2005). However, in contrast to traditional methods where the goal is to solve one large optimization problem, the MILO method is based on robustly
solving a sequence of small local (nonlinear least squares) subproblems, and then defining a globally dense displacement field with grid independent moving least squares interpolation applied to the local solutions. In order to provide robustness to potentially extreme deformations, both the forward (reference to target) and backward (target to reference) deformations are solved for simultaneously by applying the local-global procedure iteratively, with the previous forward and backward iterates providing the initial guess to the local optimization problems for the current iterates. This forward-backward iteration, which comprises the MILO method, ceases when the forward map and the backward map are consistent with one another.

The benefits of the DIR subproblem approach taken by the MILO method are threefold. First, the formulation of a local optimization problem incorporates noise robustness into the DIR solution by describing the unknown displacement as a local least squares fit to the voxel motion-intensity model. Second, the local approach breaks the full DIR problem into a series of smaller, more manageable subproblems, a process that easily lends itself to parallel implementations (e.g. a Graphics Processor Unit implementation). Finally, in our case, the local optimization formulation serendipitously allows us to describe the unknown $\lambda$ as a function of the displacement vector $d$, thereby keeping the number of unknowns (three at each voxel) the same as with traditional DIR methods.

In order to facilitate quantitative spatial accuracy evaluation, we first developed a reference dataset of image pairs, no less than 618 (max: 1172) anatomical landmark feature pairs were identified between image pairs, and further characterized for statistical uncertainty. The reference data were obtained using the methods previously described by Castillo et al. (Castillo et al., 2009b), and represent the only dataset of its kind for images obtained directly from the
COPDgene study. Using these expert feature pairs as reference, we found the MILO algorithm resulted in average Euclidean magnitude spatial registration errors ranging from 0.96 – 2.47 mm per case and 1.27 mm overall. As a comparison, the spatial accuracy performance of the well known and previously reported passive force demons algorithm (Gu et al., 2010; Thirion, 1998) yielding mis-registration of the landmark features on the order of the original non-registered positions (see Table 2). Previous work has shown the demons DIR algorithm to perform with spatial accuracies on the order of approximately 1-2 mm for registration of the extreme phases obtained from 4DCT. Evaluation of the MILO spatial accuracy over these same 4DCT images yielded spatial accuracies that were statistically indistinguishable from measurements of the reference manual registration error (see Table 2). This discrepancy in spatial accuracy performance between methods speaks not only to the inherent difficulties associated with deformable registration of the COPDgene image data, but also to the potential dangers of blind application of a single algorithm for multiple purposes, without careful performance evaluation that is specific to each new setting. The reference data generated in this study have been made publically available online (http://www.dir-lab.com) to facilitate precisely this type of analysis.

Other DIR algorithm advancements have addressed aspects of these issues. Inverse consistency was first introduced by Christensen et al. (Christensen and Johnson, 2001) and subsequently applied to the tracking of lung tissue using DIR registration of gated CT images from resting tidal breathing subjects (Christensen et al., 2007). In those studies, inverse consistency was incorporated into the cost function (Christensen and Johnson, 2001). In this study, MILO utilizes a consistency ansatz as an acceptance criteria for selecting solutions to localized DIR subproblems. Yin et al. (Yin et al., 2009) describe a mass preserving B-spine DIR
which performs with high spatial accuracy for BH-CT images with large displacements. In that study, the CT images were from normal volunteers, and the acquisition techniques for both the BH-CT images were nearly the same as the COPDgene higher dose inspiratory BH-CT. Their algorithm was not evaluated for the effects of COPD on the lung mechanics and the image noise. In particular, the use of B-spline representation of the image data and Newton's method both would render the algorithm susceptible to performance degradation from image noise. Ding et al. (Ding et al., 2009) evaluated whole-lung versus lobe-by-lobe DIR using high resolution BH-CT images acquired at the higher dose from the same image data and DIR algorithm as Yin et al. (Yin et al., 2009). Lobe segmentation of each phase and subsequent registration on a lobe-by-lobe basis improved the spatial accuracy of the resulting registration, 0.83 mm versus 0.73 mm, especially in regions near the lobe borders.

Further improvements to MILO would come through incorporation of additional anatomic information, such as lobe-by-lobe registration, and a GPU implementation to increase speed. Knowing when a DIR algorithm performs well or not for a given instance remains an uncertainty. Our next research step in the development of this DIR for clinical and research applications is to develop automated quality assurance measures. In routine application use, a report of these measures would accompany each DIR usage, allowing the user to accept or reject the DIR result for clinical or research use.
6. Conclusions

In this study, we introduced the Moving Least Squares Guided Local Optimization (MILO) DIR algorithm, designed to register noisy image pairs depicting drastic deformations. We tested its performance using 4378 landmark point pairs from 5 COPDgene BH-CT image pairs. MILO achieved an average spatial accuracy of 1.27 mm. As a reference for comparison, the 4DCT passive force demons algorithm achieved an average spatial accuracy of 21.37 mm.

Acknowledgements

We most sincerely thank the University of Texas M. D. Anderson Cancer Center Physician-Scientist Program who provided support for this project. This work was also partially supported by the National Cancer Institute through NIH/NCI Grant (R21CA128230) and through an NIH Training Grant (T32CA119930). JM was supported by a post-doctoral training grant, NCI Grant (R25T-CA90301). The authors also wish to acknowledge Julia Hoellenriegel who assisted in the preparation of this manuscript. The authors thank Dr. Todd Pawlicki for facilitating the collaboration with the medical physics and diagnostic radiology thoracic groups at UCSD for this study.
Vitae

Edward Castillo, Jr. PhD. - Edward Castillo currently holds a research scientist position in the Department of Radiation Oncology at the University of Texas M. D. Anderson Cancer Center. He earned his M.A. (2005) and Ph.D (2007) from the Computational and Applied Mathematics department at Rice University. His current research areas include image processing, surface reconstruction from noisy point cloud data, and deformable image registration. Dr. Castillo also holds an adjunct appointment in the Computational and Applied Mathematics Department at Rice University.

Richard Castillo, M.Sc. - Richard Castillo is currently a doctoral graduate student in the department of imaging physics at the University of Texas M. D. Anderson Cancer Center, through the UT Graduate School of Biomedical Science, where he earned his M.S. in 2007. His current research interests include quantitative performance evaluation of deformable image registration, and CT-based functional lung imaging.

Xuejun Gu, Ph.D. - Xuejun Gu is a post-doctoral scholar in the department of radiation oncology at the University of California San Diego. She completed her Ph.D. (biomedical engineering) at Columbia University in 2009. Her current research is on the use of high performance computing for real-time on-line adaptive radiotherapy. She is a member of the Center of Advanced Radiotherapy Technologies the research arm of the Department of Radiation Oncology.
Josue Martinez, Ph.D. – Dr. Maritnez earned his Ph.D. in statistics from Rice University and has completed a postdoctoral research fellowship in the bioinformatics training program at Texas A&M University. He currently holds an assistant professor appointment in the Department of Epidemiology & Biostatistics at Texas A&M University. His research interests include functional data analysis, medical imaging, and microarray data analysis.

David Fuentes, Ph.D. - David Fuentes is a post-doctoral scholar in imaging physics at the University of Texas M. D. Anderson Cancer Center. He completed his Ph.D. (computational science and engineering) from the University of Texas Austin in 2008. His current research includes deformable image registration, computational modeling, and real-time control of patient-specific laser treatment of cancer.

Steve B. Jiang, Ph.D. - Steve Jiang is a tenured associate professor at the University of California San Diego where he is the executive director of the Center for Advanced Radiotherapy Technologies. He completed his Ph.D. (medical physics) from the Medical College of Ohio in 1998 and a postdoctoral fellowship at Stanford University in 2000. From 2000 until 2007 he was an assistant professor at Massachusetts General Hospital and Harvard Medical School. Currently his research has focused on the application of innovative physics and engineering technologies to improve the effectiveness and efficiency of cancer radiotherapy.

Paul Friedman, M.D. - Professor Friedman is currently Professor Emeritus of Radiology at the University of California San Diego. He was elected to Alpha Omega Alpha prior to receiving his
M.D. from Yale University in 1960. He was elected to Pi Mu Epsilon and Phi Beta Kappa upon completing his undergraduate mathematics degree from the University of Wisconsin in 1955. Fellow in Academic Medicine of the Picker Foundation, 1966-68, studying pulmonary pathology under A.A. Liebow. Head of Chest Radiology at UCSD from 1968-2001. From 1982-95, he was the Associate Dean and Dean for Academic Affairs of the School of Medicine. He studies quantitative pulmonary CT.

Thomas Guerrero, M.D., Ph.D. - Thomas Guerrero is a tenured associate professor at the University of Texas M.D. Anderson Cancer Center. He completed his M.D. and Ph.D (biomedical physics) from the UCLA David Geffen School of Medicine in 1997 and residency training in radiation oncology at Stanford University in 2002. His research efforts are focused on deformable image registration and physiological & molecular imaging of radiation lung injury. He has an adjunct appointment in the Computational and Applied Mathematics Department at Rice University.
FIGURE LEGENDS

Figure 1. Breath-hold CT image pair from the COPDgene study. a) The inhale BH-CT shown in coronal section was acquired at high tube current (200mA) resulting in a low noise image. b) The corresponding exhale BH-CT coronal section. The eBH-CT was acquired at low tube current (50 mA) resulting in a noisy image. Note the mottled appearance in the soft tissue regions compared in the iBH-CT. c) Plot of the histogram distribution of lung CT values in Hounsfield units (HU) from the BH-CT image pair shown in "a" and "b".

Figure 2. Anatomic deformation. A vessel bifurcation is identified on both the inhale & exhale BH-CT image pair, denoted by the green cross. There is a substantial change in the shape of the structures that make-up the bifurcation. On the inhale BH-CT the blood vessels are straight and the lung tissue is dark. On the exhale BH-CT the blood vessels are bent, due to the substantially reduced volume of the lung, and the surrounding lung tissue is lighter. The bifurcation is located much closer to the lung border on the exhale BH-CT. These changes make automated selection of this point-pair using cross-correlation difficult.

Figure 3. Landmark points. Landmark point pairs were selected on exhale and inhale BH-CT images from 5 COPDgene cases. The number of pairs were 773, 618, 1172, 786, and 1029 for each cases respectively. A vector field projection is shown for each case using the color map for displacement
length (0 to 50 mm) shown at right. Those vectors with > 50 mm displacement are mapped to the same color as 50 mm displacements. The range of displacements vary for each case, reflecting the extent of their COPD disease.

**Figure 4.** DIR example results. **a)** Spatial error histogram with bins in 5 mm increments for the demons and MILO DIR algorithms is shown for case 5. **b)** Calculated displacement vectors from the MILO DIR algorithm, and **c)** the corresponding Euclidean spatial errors. **d)** Calculated displacement vectors from the demons DIR algorithm, and **e)** the corresponding Euclidean spatial errors. The mean spatial errors were 0.96 mm for the MILO DIR and 29.45 mm for the demons DIR.

**Figure 5.** DIR comparative evaluation. The box plot compares the spatial accuracy of the MILO versus demons DIR algorithms **a)** over the 5 4DCT cases, and **b)** over the 5 COPDgene cases. Note that (a) and (b) are presented on different scales due to the large difference in error ranges between 4DCT and COPDgene datasets.

**Figure 6.** Local divergence. 618 landmark point pairs were selected on the exhale and inhale BH-CT images from case 2. A vector field projection is shown for each lung using the color map for displacement length (0 to 35 mm) shown at right. Those vectors with >35 mm displacements are mapped to the same color as the 35 mm displacements. In this case, there was a local region with a
significant divergence illustrated by white lines. Interpolation in this region is prone to error, which likely caused this case to have the highest measured spatial error (2.47 mm) for the MILO DIR.
References


Table 1. **CT image and reference data characteristics.** The image and voxel dimensions are shown for all cases included in this study. Also shown are the number of spatial accuracy reference landmarks for each case, along with the corresponding average (and standard deviation) landmark displacement. Estimates of observer variance in landmark registration were obtained by repeat registration as described in Section 2.2.3, and are also shown as mean (and standard deviation), combined for the set of multiple observers. All measurements of distance are reported in units of millimeters.

<table>
<thead>
<tr>
<th>Case</th>
<th>Image Dimension</th>
<th>Voxel Dimension (mm)</th>
<th># Landmarks</th>
<th>Avg (SD) Displacement (mm)</th>
<th>Observer Error (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>4DCT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DCT-1</td>
<td>256 × 256 × 94</td>
<td>0.97 × 0.97 × 2.50</td>
<td>1280</td>
<td>4.01 (2.91)</td>
<td>0.85 (1.24)</td>
</tr>
<tr>
<td>4DCT-2</td>
<td>256 × 256 × 112</td>
<td>1.16 × 1.16 × 2.50</td>
<td>1487</td>
<td>4.65 (4.09)</td>
<td>0.70 (0.99)</td>
</tr>
<tr>
<td>4DCT-3</td>
<td>256 × 256 × 104</td>
<td>1.15 × 1.15 × 2.50</td>
<td>1561</td>
<td>6.73 (4.21)</td>
<td>0.77 (1.01)</td>
</tr>
<tr>
<td>4DCT-4</td>
<td>256 × 256 × 99</td>
<td>1.13 × 1.13 × 2.50</td>
<td>1166</td>
<td>9.42 (4.81)</td>
<td>1.13 (1.27)</td>
</tr>
<tr>
<td>4DCT-5</td>
<td>256 × 256 × 106</td>
<td>1.10 × 1.10 × 2.50</td>
<td>1268</td>
<td>7.10 (5.14)</td>
<td>0.92 (1.16)</td>
</tr>
<tr>
<td><strong>COPDgene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD-1</td>
<td>512 × 512 × 121</td>
<td>0.625 × 0.625 × 2.50</td>
<td>773</td>
<td>25.90 (11.57)</td>
<td>0.65 (0.73)</td>
</tr>
<tr>
<td>COPD-2</td>
<td>512 × 512 × 102</td>
<td>0.645 × 0.645 × 2.50</td>
<td>618</td>
<td>21.85 (6.58)</td>
<td>1.06 (1.51)</td>
</tr>
<tr>
<td>COPD-3</td>
<td>512 × 512 × 126</td>
<td>0.652 × 0.652 × 2.50</td>
<td>1172</td>
<td>12.29 (6.39)</td>
<td>0.58 (0.87)</td>
</tr>
<tr>
<td>COPD-4</td>
<td>512 × 512 × 126</td>
<td>0.590 × 0.590 × 2.50</td>
<td>786</td>
<td>30.90 (13.49)</td>
<td>0.71 (0.96)</td>
</tr>
<tr>
<td>COPD-5</td>
<td>512 × 512 × 131</td>
<td>0.647 × 0.647 × 2.50</td>
<td>1029</td>
<td>30.90 (14.05)</td>
<td>0.65 (0.87)</td>
</tr>
</tbody>
</table>
Table 2. **DIR error summaries.** Mean (and standard deviation) spatial registration errors are shown for all cases, and both algorithms. Note that an asterisk (*) indicates those cases for which the measured DIR spatial errors are statistically indistinguishable from the estimates of observer variance obtained by repeat registration.

<table>
<thead>
<tr>
<th>Case</th>
<th>Algorithm</th>
<th>RL (mm)</th>
<th>AP (mm)</th>
<th>SI (mm)</th>
<th>3D (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>4DCT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DCT-1</td>
<td>Demons</td>
<td>0.35 (0.50)</td>
<td>0.37 (0.54)</td>
<td>0.61 (1.11)</td>
<td>1.11 (1.09)</td>
</tr>
<tr>
<td></td>
<td>MILO</td>
<td>0.27 (0.47)</td>
<td>0.28 (0.46)</td>
<td>0.37 (0.90)</td>
<td>0.78 (0.96)*</td>
</tr>
<tr>
<td></td>
<td>No DIR</td>
<td>0.58 (0.62)</td>
<td>0.67 (0.79)</td>
<td>3.68 (3.04)</td>
<td>4.01 (2.91)</td>
</tr>
<tr>
<td>4DCT-2</td>
<td>Demons</td>
<td>0.35 (0.59)</td>
<td>0.37 (0.63)</td>
<td>0.53 (1.06)</td>
<td>1.04 (1.15)</td>
</tr>
<tr>
<td></td>
<td>MILO</td>
<td>0.27 (0.53)</td>
<td>0.26 (0.54)</td>
<td>0.27 (0.78)</td>
<td>0.69 (0.96)*</td>
</tr>
<tr>
<td></td>
<td>No DIR</td>
<td>0.73 (0.85)</td>
<td>0.72 (0.88)</td>
<td>4.09 (4.37)</td>
<td>4.65 (4.09)</td>
</tr>
<tr>
<td>4DCT-3</td>
<td>Demons</td>
<td>0.46 (0.64)</td>
<td>0.50 (0.67)</td>
<td>0.74 (1.20)</td>
<td>1.36 (1.20)</td>
</tr>
<tr>
<td></td>
<td>MILO</td>
<td>0.34 (0.57)</td>
<td>0.33 (0.57)</td>
<td>0.36 (0.89)</td>
<td>0.85 (1.03)*</td>
</tr>
<tr>
<td></td>
<td>No DIR</td>
<td>1.17 (1.05)</td>
<td>1.28 (1.23)</td>
<td>6.10 (4.49)</td>
<td>6.73 (4.21)</td>
</tr>
<tr>
<td>4DCT-4</td>
<td>Demons</td>
<td>0.82 (0.98)</td>
<td>1.05 (1.35)</td>
<td>1.53 (2.36)</td>
<td>2.51 (2.49)</td>
</tr>
<tr>
<td></td>
<td>MILO</td>
<td>0.46 (0.69)</td>
<td>0.52 (0.77)</td>
<td>0.60 (1.40)</td>
<td>1.26 (1.50)*</td>
</tr>
<tr>
<td></td>
<td>No DIR</td>
<td>0.94 (1.21)</td>
<td>1.42 (1.22)</td>
<td>8.98 (5.04)</td>
<td>9.42 (4.81)</td>
</tr>
<tr>
<td>4DCT-5</td>
<td>Demons</td>
<td>0.56 (0.77)</td>
<td>0.80 (1.03)</td>
<td>1.05 (1.65)</td>
<td>1.84 (1.74)</td>
</tr>
<tr>
<td></td>
<td>MILO</td>
<td>0.44 (0.65)</td>
<td>0.49 (0.73)</td>
<td>0.48 (1.07)</td>
<td>1.13 (1.22)*</td>
</tr>
<tr>
<td></td>
<td>No DIR</td>
<td>0.86 (0.96)</td>
<td>1.74 (1.67)</td>
<td>6.30 (5.45)</td>
<td>7.10 (5.15)</td>
</tr>
<tr>
<td>4DCT-Composite</td>
<td>Demons</td>
<td>0.50 (0.70)</td>
<td>0.60 (0.87)</td>
<td>0.86 (1.51)</td>
<td>1.53 (1.58)</td>
</tr>
<tr>
<td></td>
<td>MILO</td>
<td>0.35 (0.58)</td>
<td>0.37 (0.62)</td>
<td>0.41 (1.01)</td>
<td>0.92 (1.14)*</td>
</tr>
<tr>
<td></td>
<td>No DIR</td>
<td>0.86 (0.95)</td>
<td>1.15 (1.19)</td>
<td>5.73 (4.53)</td>
<td>6.29 (4.28)</td>
</tr>
</tbody>
</table>

*COPDgene*
<table>
<thead>
<tr>
<th>COPD</th>
<th>Demons</th>
<th>MILO</th>
<th>No DIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPD-1</td>
<td>4.25 (2.95)</td>
<td>22.08 (11.34)</td>
<td>8.56 (6.75)</td>
</tr>
<tr>
<td></td>
<td>0.40 (0.58)</td>
<td>0.80 (2.01)</td>
<td>0.62 (1.49)</td>
</tr>
<tr>
<td></td>
<td><strong>4.31 (3.19)</strong></td>
<td><strong>23.66 (11.27)</strong></td>
<td><strong>7.39 (5.88)</strong></td>
</tr>
<tr>
<td>COPD-2</td>
<td>3.61 (2.65)</td>
<td>18.07 (7.45)</td>
<td>11.78 (8.73)</td>
</tr>
<tr>
<td></td>
<td>0.77 (1.22)</td>
<td>1.24 (2.12)</td>
<td>1.42 (3.07)</td>
</tr>
<tr>
<td></td>
<td><strong>3.35 (2.46)</strong></td>
<td><strong>19.68 (6.92)</strong></td>
<td><strong>6.22 (5.44)</strong></td>
</tr>
<tr>
<td>COPD-3</td>
<td>1.77 (1.73)</td>
<td>4.88 (4.26)</td>
<td>4.00 (4.02)</td>
</tr>
<tr>
<td></td>
<td>0.38 (0.55)</td>
<td>0.44 (0.63)</td>
<td>0.40 (0.94)</td>
</tr>
<tr>
<td></td>
<td><strong>2.11 (1.67)</strong></td>
<td><strong>6.11 (3.50)</strong></td>
<td><strong>9.70 (6.41)</strong></td>
</tr>
<tr>
<td>COPD-4</td>
<td>6.03 (4.00)</td>
<td>16.80 (10.22)</td>
<td>16.43 (11.47)</td>
</tr>
<tr>
<td></td>
<td>0.42 (0.51)</td>
<td>0.50 (0.63)</td>
<td>0.41 (1.17)</td>
</tr>
<tr>
<td></td>
<td><strong>5.75 (4.39)</strong></td>
<td><strong>17.67 (10.25)</strong></td>
<td><strong>21.20 (14.75)</strong></td>
</tr>
<tr>
<td>COPD-5</td>
<td>4.52 (3.59)</td>
<td>22.77 (11.48)</td>
<td>15.38 (10.81)</td>
</tr>
<tr>
<td></td>
<td>0.39 (0.53)</td>
<td>0.47 (0.76)</td>
<td>0.35 (0.88)</td>
</tr>
<tr>
<td></td>
<td><strong>4.46 (3.44)</strong></td>
<td><strong>24.48 (11.31)</strong></td>
<td><strong>15.83 (11.94)</strong></td>
</tr>
<tr>
<td>COPD-Composite</td>
<td>3.88 (3.04)</td>
<td>16.12 (9.23)</td>
<td>10.81 (8.61)</td>
</tr>
<tr>
<td></td>
<td>0.45 (0.68)</td>
<td>0.63 (1.29)</td>
<td>0.57 (1.55)</td>
</tr>
<tr>
<td></td>
<td><strong>3.88 (3.10)</strong></td>
<td><strong>17.52 (9.02)</strong></td>
<td><strong>12.31 (9.69)</strong></td>
</tr>
</tbody>
</table>
### Table 3. Wilcoxon rank sum test summary.

*p*-values are shown indicating the significance of differences in mean spatial registration errors between MILO and demons algorithms, and between MILO and observer measurements, separately for 4DCT and COPDgene datasets. The statistical tests indicate significantly (*p*-value < 0.05) lower spatial errors for MILO relative to demons DIR for both classes of patient images. Furthermore, for registration of the thoracic 4DCT images, measured spatial errors were statistically indistinguishable (*p*-value ≥ 0.17) between MILO and the combined set of human observer measurements.

<table>
<thead>
<tr>
<th>Case</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
<th>Case 5</th>
<th>All Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COPDgene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[MILO, Demons]</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
</tr>
<tr>
<td>[MILO, Observers]</td>
<td>0.0013</td>
<td>0.0025</td>
<td>0.00029</td>
<td>7.8 x10^{-5}</td>
<td>0.0025</td>
<td>&lt; 1 x10^{-6}</td>
</tr>
<tr>
<td><strong>4DCT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[MILO, Demons]</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
<td>&lt; 1 x10^{-6}</td>
</tr>
<tr>
<td>[MILO, Observers]</td>
<td>0.17</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.63</td>
</tr>
</tbody>
</table>
Figure 1. **Breath-hold CT image pair from the COPDgene study.** a) The inhale BH-CT shown in coronal section was acquired at high tube current (200mA) resulting in a low noise image. b) The corresponding exhale BH-CT coronal section. The eBH-CT was acquired at low tube current (50 mA) resulting in a noisy image. Note the mottled appearance in the soft tissue regions compared in the iBH-CT. c) Plot of the histogram distribution of lung CT values in Hounsfield units (HU) from the BH-CT image pair shown in "a" and "b".
**Figure 2. Anatomic deformation.** A vessel bifurcation is identified on both the inhale & exhale BH-CT image pair, denoted by the green cross. There is a substantial change in the shape of the structures that make-up the bifurcation. On the inhale BH-CT the blood vessels are straight and the lung tissue is dark. On the exhale BH-CT the blood vessels are bent, due to the substantially reduced volume of the lung, and the surrounding lung tissue is lighter. The bifurcation is located much closer to the lung border on the exhale BH-CT. These changes make automated selection of this point-pair using cross-correlation difficult.
Figure 3. Landmark points. Landmark point pairs were selected on exhale and inhale BH-CT images from 5 COPDgene cases. The number of pairs were 773, 618, 1172, 786, and 1029 for each cases respectively. A vector field projection is shown for each case using the color map for displacement length (0 to 50 mm) shown at right. Those vectors with > 50 mm displacement are mapped to the same color as 50 mm displacements. The range of displacements vary for each case, reflecting the extent of their COPD disease.
Figure 4. DIR example results. a) Spatial error histogram with bins in 5 mm increments for the demons and MILO DIR algorithms shown for case 5. b) Calculated displacement vectors from the MILO DIR algorithm and c) the corresponding Euclidean spatial errors. d) Calculated displacement vectors from the demons DIR algorithm and e) the corresponding spatial errors. The mean spatial errors were 1.13 mm for the MILO DIR and 29.5 mm for the demons DIR.
Figure 5. DIR comparison results. The box plot compares the spatial accuracy of the MILO versus demons DIR algorithms a) over the 5 4D CT cases and b) over the 5 COPDgene cases. Note that (a) and (b) are presented on different scales due to the large difference in error ranges.
Figure 6. Local divergence. 618 landmark point pairs were selected on the exhale and inhale BH-CT images from cases 2. A vector field projection is shown for each lung using the color map for displacement length (0 to 35 mm) shown at right. Those vectors with > 35 mm displacement are mapped to the same color as 35 mm displacements. In this case, there was a local region with a significant divergence illustrated by the white lines. Interpolation in this region is prone to error, which caused this case to have the highest measured spatial error (2.7 mm) for the MILO DIR.