Evaluation of Deformable Image Registration for Lung Motion Estimation using Hyperpolarized Gas Tagging MRI

by

Qijie Huang

Graduate Program in
Medical Physics
Duke University

Date:_____

Approved:

_____
Jing Cai, Supervisor

_____
Fang-Fang Yin, Chair

_____
Bastiaan Driehuys

Thesis submitted in partial fulfillment of
the requirements for the degree Master of Science in the
Graduate Program in Medical Physics
in the Graduate School of
Duke University

2014

ABSTRACT

Evaluation of Deformable Image Registration for Lung Motion Estimation using

Hyperpolarized Gas Tagging MRI

by

Qijie Huang

Graduate Program in
Medical Physics
Duke University

Date:_____

Approved:

_____

Jing Cai, Supervisor

_____

Fang-Fang Yin, Chair

_____

Bastiaan Driehuys

An abstract of a thesissubmitted in partial
fulfillment of the requirements for the degree
of Master of Science in the
Graduate Program in Medical Physics
in the Graduate School of
Duke University

2014

Abstract

Purpose: Hyperpolarized gas (HP) tagging MRI is a novel imaging technique for direct measurement of lung motion during breathing. This study aims to quantitatively evaluate the accuracy of deformable image registration (DIR) in lung motion estimation using HP tagging MRI as references.

Method: Three healthy subjects were imaged using the HP MR tagging, as well as a high-resolution 3D proton MR sequence (TrueFISP) at the end-of-inhalation (EOI) and the end-of-exhalation (EOE). Ground truth of lung motion and corresponding displacement vector field (tDVF) was derived from HP tagging MRI by manually tracking the displacement of tagging grids between EOI and EOE. Seven different DIR methods were applied to the high-resolution TrueFISP MR images (EOI and EOE) to generate the DIR-based DVFs (dDVF). The DIR methods include Velocity (VEL), MIM, Mirada, multi-grid B-spline from Elastix (MGB) and 3 other algorithms from DIRART toolbox (Double Force Demons (DFD), Improved Lucas-Kanade (ILK), and Iterative Optical Flow (IOF)). All registrations were performed by independent experts. Target registration error (TRE) was calculated as tDVF – dDVF. Analysis was performed for the entire lungs, and separately for the upper and lower lungs.

Results: Significant differences between tDVF and dDVF were observed. Besides the DFD and IOF algorithms, all other dDVFs showed similarity in deformation magnitude

distribution but away from the ground truth. The average TRE for entire lung ranged 2.5-23.7mm (mean=8.8mm), depending on the DIR method and subject's breathing amplitude. Larger TRE (13.3-23.7mm) was found in subject with larger breathing amplitude of 45.6mm. TRE was greater in lower lung (2.5-33.9 mm, mean=12.4mm) than that in upper lung (2.5-11.9 mm, mean=5.8mm).

Conclusion: Significant differences were observed in lung motion estimation between the HP gas tagging MRI method and the DIR methods, especially when lung motion is large. Large variation among different DIR methods was also observed.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my deepest appreciation to my advisor, Dr. Jing Cai, for his time and guidance throughout this project and Dr. Fang-Fang Yin and Dr. Bastiaan Driehuys for serving on my committee and offering their advice and expertise.

I would also like to acknowledge MIM Software, Mirada Medical Inc, and You Zhang for performing image registration for my project. Additionally, I would like to thank Yilin Liu and Lei Hu as well for their assistant with data analysis.

1. Introduction

1.1 Overview of Deformable Image Registration (DIR)

Image-guided radiation therapy which involves frequent or daily image acquisition has been widely used in clinic for target delineating and patient positioning. In the planning and setup process, these images acquired on board are usually mapped with the simulation images using rigid registration. [1] However patient's anatomy and position might be changing during the whole treatment process, especially for breathing motion in lung and tumor shrinkage as a response to radiation therapy. Thus the deformable image registration (DIR) is desirable for more accurate patient care taking into account these anatomical and biological variations.

The goal of deformable image registration (DIR) is to find the transformation between two images so that the differences in these two images will be minimized. This transformation generated from images can be applied to the contours of the organs and target as well as the dose distribution calculation. Nowadays DIR has many potential applications in radiation therapy research from automatic time propagation of region of interest (ROI) in adaptive radiation therapy (ART) [2] to 4D breathing motion modeling [3].

In ART, the DIR is used when matching each voxel in daily images with their corresponding points in planning images. As a result, new contours for organs at risk

(OAR) and the target can be delineated. Then the dose distribution can be computed with irradiation parameters. The dose accumulation for the OAR and the target will be calculated within the new contour.

A 4D thorax model allows evaluating the motion and displacement of the lung quantitatively as a function of time. The model is generated from a series of 3D images acquired over a continuous breathing cycle. By registering these 3D images and recording the corresponding deformation maps, each voxel in the lung will be given a mathematic function over time.

These applications all require precise deformable image registration. Small errors in the deformation field near the target would result in large variation especially in regions with sharper dose fall off. Before the DIR algorithms are to be implemented in clinical use, quantitative evaluation of DIR algorithms' accuracy is needed.

1.2 Challenges in DIR Validation

Qualitative evaluation of deformable image registration is usually accomplished by visual evaluation. Visual evaluation for deformable image registration is convenient and common. Usually after applying a registration, a blending of deformed image and the target image are available through the DIR software. Users will be able to compare the two images by looking back and forth in key regions of interest. However, this visual

evaluation can be subjective and varies from person to person. A more objective and quantitative evaluation tool is needed.

Current quantitative DIR evaluation methods that are used include estimation of registration errors for landmarks or contours [4-7]. A lot of research has been involved with physical phantoms with embedded landmarks. The deformation of these phantoms is known and the landmarks inside the phantom are easy to identify from the images acquired. Kirby et al proposed a landmark-based evaluation with a developed phantom which represents an axial plane of pelvic anatomy. [8] Different parts of the organs were built with different materials to simulate according tissue. Markers in the phantom were measured during the deformation and this indicated the actual displacement. A comparison of the displacement of markers from the deformation map with the actual displacement was analyzed afterwards. However in reality, the patients' anatomy and motion can be more complex with geometrical and biological uncertainties. Lafiti et al proposed an evaluation method on lung patients based on anatomical landmarks from patients' 4D CT images. These anatomical landmarks were manually delineated by radiologists based on various features in the lung. [9] The problem and ambiguity with this method is that the number of landmarks that can be identified is usually not sufficient enough to represent the accuracy of deformation in the entire region. Moreover, the landmarks are not always available in homogenous regions. [10]

Therefore, a quantitative and objective assessment of DIR is necessary. A desired evaluation method would be based on real patient anatomy to ensure that it's clinical applicable, in which the number of points for validation is large enough to demonstrate the deformation in the whole region.

1.3 Aim of This Study

The aim of this study is to develop and test a methodology for quantitative evaluation of DIR in the thorax using HP gas tagging MRI in a pilot human study.This method allows for direct measurement of lung deformation during breathing, providing a physiological ground truth for evaluating DIR of the lungs. We will compare lung deformation measurements between the HP tagging MRI method and various DIR methods, and investigate impacting factors that affect the accuracy of DIR of the lungs.

2. Material and Methods

2.1 Study Design

We proposed a DIR evaluation method using direct in vivo lung motion measurement from hyperpolarized gas tagging MR as reference. There are mainly two parts in the evaluation process. The first step is to generate deformation vector field (DVF) for the reference and evaluated DIR methods. Next is to perform global and regional quantitative comparison between reference DVF and evaluated DVFs.

Ground truth of lung motion and corresponding displacement vector field (tDVF) were derived from HP He-3 tagging MRI by manually tracking the displacement of tagging grids between end-of-inhalation (EOI) and end-of-exhalation (EOE) respiratory phases.

Seven different DIR methods were applied to the high-resolution MR images (2.5mm x 2.5 mm x2.5 mm, acquired with Siemens in TrueFISP sequence) images (EOI and EOE) to generate the DIR-based DVFs (dDVF).

The general work flow chart is illustrated in Figure 1.

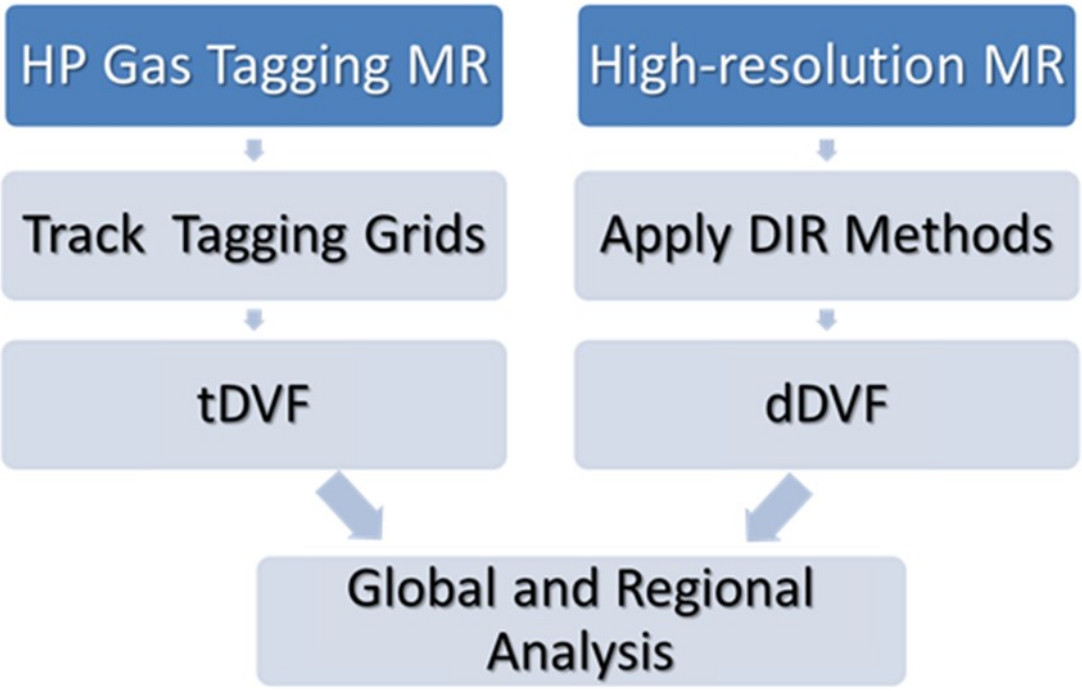

Figure 1 A work flow for evaluation of DIR methods.

2.2 Ground Truth DVF from Hyperpolarized Gas Tagging MRI technique

This study uses a novel hyperpolarized gas tagging MR technique as reference for DIR method validation. Proton (1H) MR tagging is a technique that generates spatially encoded pattern on tissue. The proton MR grid-tagging technique applies sinc-modulated RF-pulse trains along the frequency- and phase-encoding directions followed by a multi-slice fast slow-angle shot (FLASH) –based acquisition. [11] This technique has been widely used for evaluation of motion and displacement of myocardial tissue. However, in the lung area, due to the low proton intensity and numerous magnetic-

susceptibility interfaces, the quality of direct grid-tagging lung images is not ideal enough for quantitative assessment.

A grid-tagging technique using HP He-3 as gaseous contrast was developed. [12, 13] The hyperpolarized gas, when inhaled, provides a very high signal in the lung airspace. When combined with the grid-tagging technique, lung images with bright clear tagging grids can be acquired. If we acquire the lung image at two breathing phases, EOI and lung respiratory motion can be then measured directly by tracking the displacement of tagging grids.

As shown in Figure 2, the tagging grids in the lung area are clearly visible in both the EOI phase and the EOE phase. By tracking the movements of these tagging grids between the two breathing phases, the displacement of corresponding lung area can be measured directly in vivo.

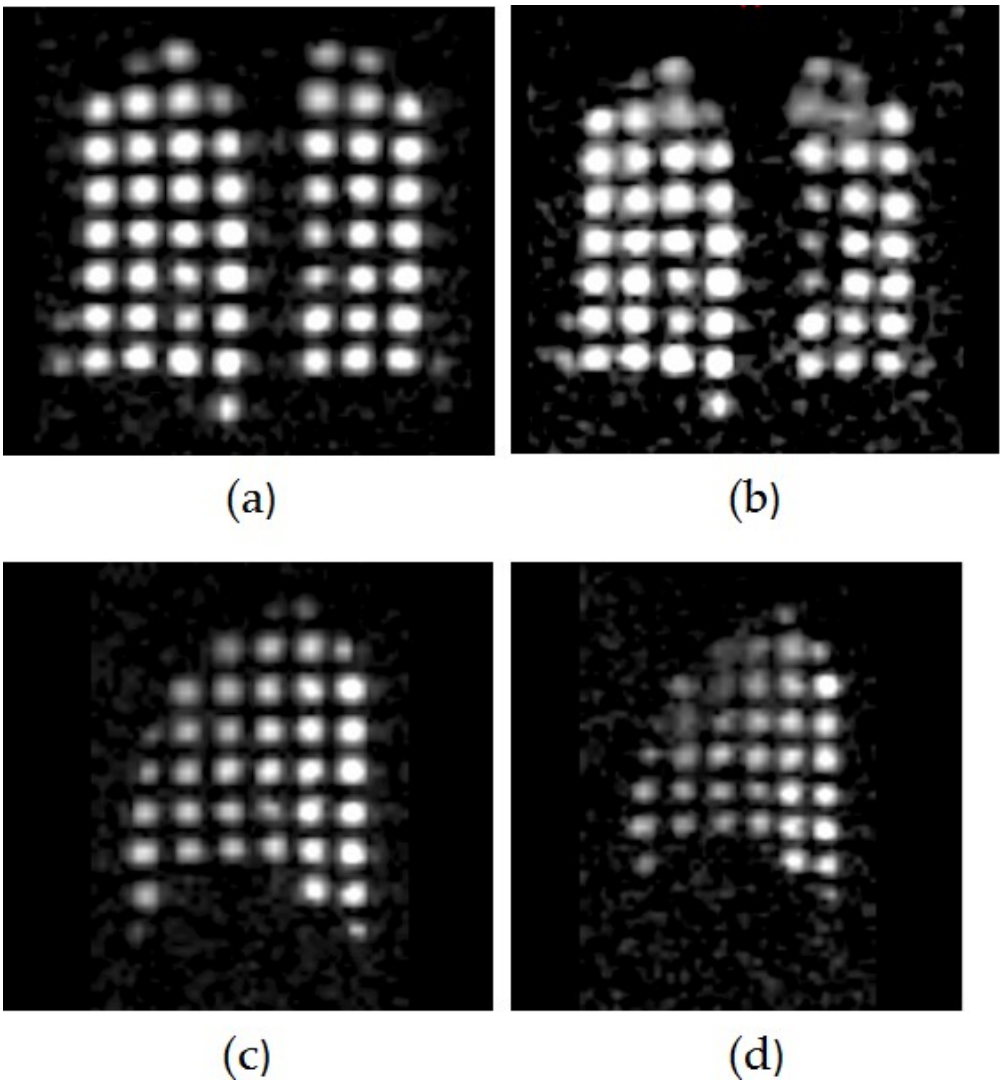

Figure 2 Examples of HP He-3 MR images in coronal view at (a) EOI and (b) EOE, and in the sagittal view t (c) EOI and (d) EOE.

2.3 Deformation Image Registration Methods Evaluated

Seven different DIR methods were evaluated. The DIR methods include Velocity (VEL), MIM, Mirada, multi-grid B-spline from Elastix (MGB) [14, 15] and 3 other algorithms from DIRART toolbox [16] (Double Force Demons (DFD), Improved Lucas-Kanade (ILK), and Iterative Optical Flow (IOF)).

All the DIR methods were performed separately by experts. The EOI high-resolution proton MR images and EOE high-resolution proton MR served as target and moving images, respectively.

Three commercial software package were evaluated, Velocity AI (v 3.0.1) (Velocity Medical, Atlanta, GA), MIM (v 6.0) (MIM Software Inc., Cleveland, OH) and Mirada Medical RT Workstation (Mirada Medical Limited, England). The DIRART toolbox and the Elastix are both open sources DIR tools that are implemented into MATLAB (The MathWorks, Natick, MA). For all the DIR methods, the deformed images were first visually evaluated by overlaying deformed images with fixed image to ensure reasonable accuracy and then the DVFs were exported from the software for further quantitative analysis.

For the commercial software, information on the details of parameter numbers used in the software is not provided by the vendor, nor can the user change parameter settings for deformation. However, an optimized deformation work flow could still increase the deformation accuracy. Multipass B-spline method is used in Veloity AI. MIM software utilizes free form deformable registration method. In Mirada Medical, radial basis function is used for transformation.

The DIRART tool provides a large number of adjustable parameters that would vary DIR performance. The desirable parameters could be chosen based on a known ground truth. However, here we would like to evaluate the performance of DIR as

general, default parameters are utilized hence. Three different algorithms were chosen for validation. These algorithms provide visually reasonable deformation results.

DIR process involves three main components: deformation model, matching criteria and optimization method [17]. The seven DIR methods evaluated are classified by different deformation models. A summary of DIR methods is shown in Table 1.

Table 1 Summary of DIR methods evaluated.

Deformation Models		DIR Methods
Geometric transformations derived from physical models	Diffusion Models	DFD
	Viscous Fluid Flow Models	IOF
		ILK
Geometric transformations derived from interpolation theory	Radial basis functions	Mirada
	Free-form deformation	MIM
		VEL
		MGB

2.4 DVF Generation

2.4.1 DVF

For two images M and F, while M is the moving image and F is the fixed image. Deformable image registration is to compute the deformation vector field DVF in order to optimize the system energy equation:

$$E = \int_{\Omega} S(DVF(M), F)d\Omega + \alpha^2 \int_{\Omega} R(V)d\Omega$$

The DVF is usually 'pull-back' vector field [18], meaning the DVF is defined on the coordinates of target image. It has the same dimension as target image and each element of DVF is a 3D vector, associated with the voxel displacement in 3 directions. The equation for definition can be described as.

$$F(i, j, k) = M(i + \Delta i, j + \Delta j, k + \Delta k)$$

$$DVF(i, j, k) = \begin{pmatrix} \Delta i \\ \Delta j \\ \Delta k \end{pmatrix}$$

Where the i, j, k denote the voxel position of point N, and the $\Delta i, \Delta j, \Delta z$ denotes the 3D vectors of DVF for the point respectively.

2.4.2 Image Acquisition

Three healthy subjects were imaged at the EOI and EOE phasesin a single scan. For each subject three image data sets were acquired in total, including a HP He-3 tagging MR, a low-resolution 3D proton MR sequence (TrueFISP) as well as a high-resolution 3D proton MR sequence (TrueFISP). The low-resolution proton MR and the tagging MR are acquired at the same time with same breath hold. While the high-resolution proton MR is acquired afterwards in two separate breath holds to capture richer internal lung features (pulmonary vessels). The breathing amplitude was

monitored by subject inhaling 1.5 L gas and exhaling 0.75L gas for both breath holds to ensure two breath holds were same.

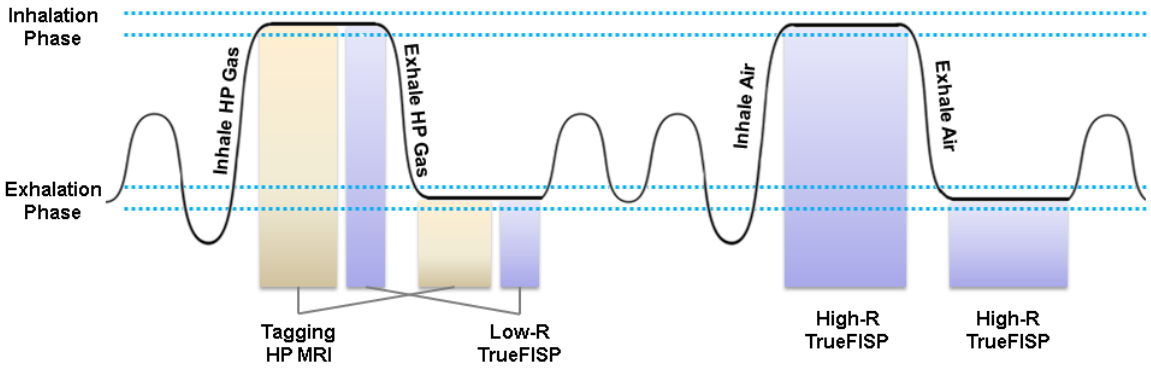

Figure 3 Image acquisition workflow

The image resolution for HP He-3 tagging MR and low-resolution 3D proton MR images are both 0.45 cm in all directions, and 0.25 cm for high-resolution 3D proton MR images.A tagging grid size is the spatial diameter of the grid point. We also defined the grid resolution as the distance between adjacent tagging grid centers. The tagging grid point size is about 1.50 cm in 3 directions. And the tagging grid resolution is 2.00 cm, indicating that the distance between adjacent grid centers is 2.00cm.

Given that HP He-3 tagging MR can provide information of direct measurements of lung motion through tagging grids, as we can see from Figure 1, the absolute spatial position for these tagging grids, however, remains unknown due to the lack of information on other organs in subjects. The rationale for acquiring low-resolution 3D proton MR is that the low-resolution proton MR can work as an indicator for identify spatial position of these grids. The low-resolution 3D proton MR and the tagging MR

share the same spatial coordinates, but the low-resolution proton MR carries detailed information on the entire body.

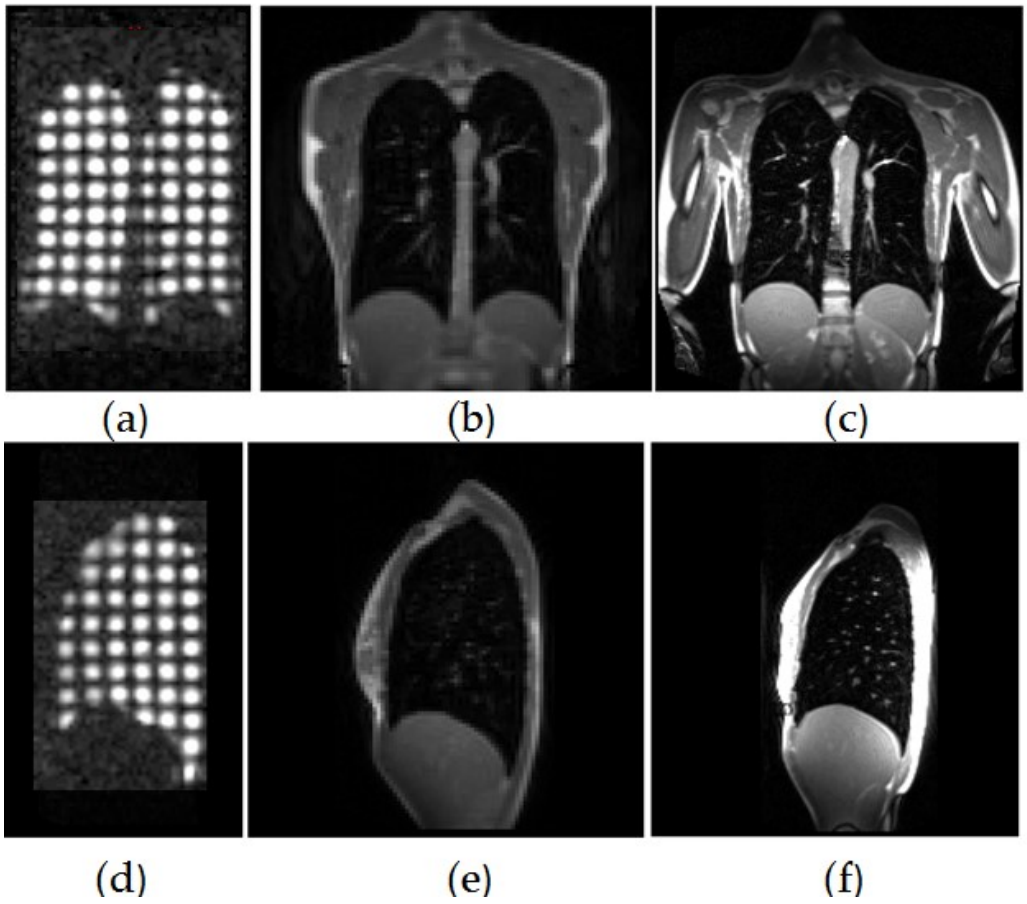

Figure 4 Top are examples of MR images in coronal view for (a) HP He-3 Tagging MR, (b) low-resolution proton MR, (c) high-resolution proton MR. Bottom are MR images in sagittal view for (d) HP He-3 Tagging MR, (e) low-resolution proton MR, (f) high-resolution proton MR.

Figure 4 illustrates the three images data sets in coronal and sagittal views. The spatial position for tagging grids is measured according to low-resolution MR images. A rigid registration is performed between low-resolution proton MR and high-resolution

proton MR images. This rigid registration mainly accounts for coordination changes between these two proton MR images.

As we can see from Figure 4, fine details of lung structures are clearly visible from the high-resolution proton MR images. The high-resolution proton MR images were used to evaluate the DIR methods.

Thus we acquired 3 image data sets, the HP He-3 tagging MR and the low-resolution proton MR images for direct in vivo lung motion measurement and tagging grid position indication, as well as high-resolution proton MR images for deformable images registration using different DIR methods.

2.4.3 DVF Generation Process

Ground truth of lung motion and corresponding displacement vector field (tDVF) were derived from HP He-3 tagging MRI by manually tracking the displacement of tagging grids between EOI and EOE phases. Seven different DIR methods were applied to the high-resolution TrueFISP MR images (EOI and EOE) to generate the DIR-based DVFs (dDVF). The DIR methods include Velocity (VEL), MIM, Mirada, multi-grid B-spline from Elastics (MGB) [9-10] and 3 other algorithms from DIRART toolbox [11] (Double Force Demons (DFD), Improved Lucas-Kanade (ILK), and Iterative Optical Flow (IOF)).

For the tagging MR images, usually around 300 tagging grids were observed clearly for evaluation. Spatial position of centroid of each tagging grid was recorded for

EOI and EOE phases respectively as P_{EOI}, P_{EOE}. A special coordinate for these tagging grids is also generated to describe the relative tagging grids positions in the ML, AP and SI directions. This specific coordinate is created to facilitate identification of corresponding tagging grids between the EOI and EOE phases. After the spatial positions of all tagging grids are identified, the ground truth DVF was calculated as: .

$$tDVF = P_{EOE} - P_{EOI}$$

The displacement vector field is assigned to the central voxel of each HP He-3 tagging grid. It should be noted that due to the subject exhalation, not all the grids observed in EOI phase could be tracked in EOE phase. Some grids information has to be discarded because of the ambiguity of position identification. The tagging grids recorded here for evaluation purpose are the ones that demonstrated clear boundary in both EOI and EOE phases.

Different DIR methods were applied to the high-resolution proton MR to generate the DVF. The high-resolution proton MR images (iso-voxel of 2.5mm) demonstrated rich pulmonary vascular structures, which are important for generating accurate DIR-based DVFs. For all DIR methods, the high-resolution proton MR images at EOE phase were registered to the EOI phase, designating the EOI phase as the target. The DVFs generated had the same size as the high-resolution proton MR images..

The tDVF from tagging is a sparse DVF, where only the pixel at center of grid points contains a DVF vector. On the other hand, DIR-derived DVFs have the same

resolution as the original MR images, with DVF calculated for each voxel in the proton MR images. To quantitatively compare the two DVFs, a down sampling of DIR-derived dDVF was performed.

Since the tagging grid displacement represents the average displacement of points within grid mass sphere, DVF from DIR methods were thus down sampled by averaging DVF in the corresponding tagging grid areas.

$$\mathrm{dDVF}(x_j, y_j, z_j) = \frac{1}{N}\sum_{i=1}^{N} DVF(x_i, y_i, z_i)$$

Where the DVF denotes the displacement vector, and x_i, y_i, z_i represent the location of the *i*th pixel within the according tagging grid j (total of N pixels). x_j, y_j, z_j denotes the center pixel location of the *j*th tagging grid.

This down-sampled dDVF was also assigned to the pixel as indicated as the center of the grid mass.

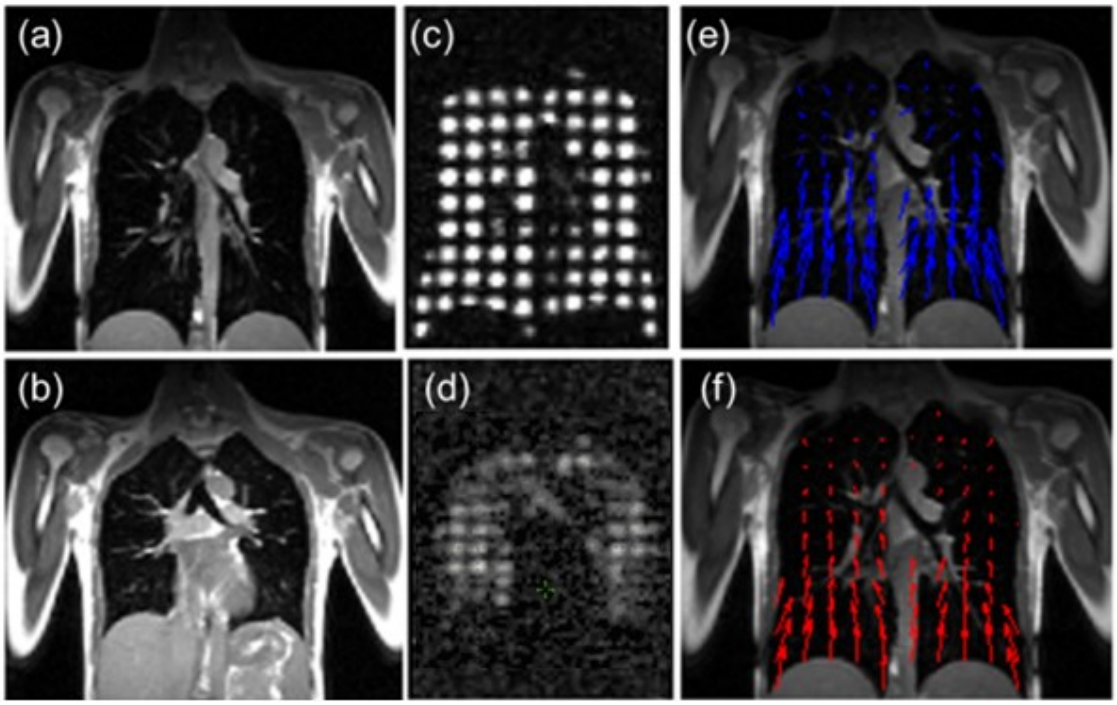

Figure 5 Examples of MR images and DVFs in the coronal view. High-resolution TrueFISP proton MR images at (a) EOI and (b) EOE; HP He-3 tagging MR images at (c) EOI and (d) EOE; and the corresponding (e) tDVF and (f) dDVF (from VEL).

An example of tDVF and down sampled dDVF as well as proton MR and tagging MR in EOI and EOE is illustrated in Figure 5. The tail of the arrow demonstrates the grid center in EOI phase, while the arrow points to its corresponding EOE phase position.

2.5 Evaluation Methods

We evaluated the deformable image registration methods with direct measurement of lung motion. The variable we look into is the DVF generated from ground truth and the DIR methods.

2.5.1 TRE Comparison: Overall Statistics and Frequency Distribution

Target registration error (TRE) was calculated as:

$$TRE=tDVF-dDVF$$

Evaluation of TRE is performed globally for the entire lung, as well as regionally for the upper and lower lung, respectively. The TRE was calculated and compared in three dimensions, medial-lateral (ML) anterior-posterior (AP) and superior-inferior (SI). Comparison in the magnitude of 3D DVF was also performed. The global accuracy for each registration method was demonstrated by calculating the mean TRE, standard deviation and the maximum TRE over all tagging grids for each subject.

Error frequency maps were also plotted for each subject, illustrating the error distribution in the amplitude of 3D DVF.

2.5.2 DVF Magnitude Comparison-Distribution and Correlation

The metrics of mean TRE and standard deviation provide a basic understanding of performance for each registration method. However a more comprehensive evaluation method is desired for demonstrating error distribution for each tagging grids. Thus, we proposed a DVF magnitude distribution map calculated for each subject.

The DVF magnitude for each grid was divided into bins of 2 mm. A distribution map illustrates the number of grids that demonstrated the corresponding displacement within the bin. In this way, an overall comparison of displacement can be achieved.

Based on the DVF magnitude distribution, a correlation study between different DIR methods and the ground truth is also conducted and presented.

2.5.3 Spatial Error Distribution

Neither the mean metric nor the DVF magnitude distribution could reveal the spatial error information which indicates where the error exists. So here we also presented a TRE color map in coronal view for each DIR method in ML and SI directions. The colors in different lung regions stand for magnitude of error according to the color bar. Error distribution spatially is easily interpreted from the TRE color map.

2.5.4 A Correlation Study on TRE and Lung Motion

It is of interest to investigate the impacting factors of the TRE. In this study we evaluated the correlation between lung motion and TRE. The lung motion is defined as the ground truth 3D motion calculated from tagging MR.

3. Result

3.1 TRE Comparison: Overall Statistics and Frequency Distribution

This section compares the TRE between different DIR methods. Firstly, an overall statistics including mean TRE, standard-deviation and maximum TRE were calculated. Secondly, a 3D error frequency distribution maps were plotted for seven DIR methods in three subjects.

3.1.1 Overall TRE Statistics

Table 2 summarizes of the results of TRE of the entire lung for the three subjects. In general, large TRE (14.2 - 22.5 mm) was observed in Subject 1. Smaller TRE was observed in Subject 2 and Subject 3(2.5 mm-6.4mm, 4.4 mm-5mm, respectively).

Table 2 Target Registration Error (in mm) for entire lung in 3 healthy subjects.

		ML	AP	SI	Magnitude
		Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
Sub #1	VEL	4.0 ± 3.6	10.2 ± 6.1	9.9 ± 7.8	14.6 ± 8.0
	MIM	4.5 ± 3.5	8.4 ± 5.9	11.0 ± 8.7	14.2 ± 8.3
	DFD	4.3 ± 3.4	16.2 ± 10.6	18.4 ± 15.3	22.5 ± 15.1
	ILK	4.5 ± 3.9	18.3 ± 11.3	18.6 ± 16.0	22.4 ± 16.5
	IOF	4.5 ± 4.1	17.7 ± 11.6	17.7 ± 14.3	23.7 ± 15.4
	MBS	5.5 ± 4.7	12.0 ± 8.8	13.8 ± 11.6	16.0 ± 9.8
	Mirada	6.7 ± 4.8	15.0 ± 8.8	17.9 ± 13.2	21.3 ± 13.1
Sub #2	VEL	2.4 ± 1.9	2.2 ± 1.6	2.1 ± 1.6	2.5 ± 1.7
	MIM	3.1 ± 3.1	3.0 ± 2.7	3.2 ± 2.7	3.2 ± 2.5
	DFD	4.8 ± 2.2	4.0 ± 2.7	7.9 ± 2.7	6.4 ± 3.4
	ILK	2.0 ± 1.7	5.7 ± 3.8	6.6 ± 4.2	3.0 ± 2.2
	IOF	3.2 ± 1.6	4.5 ± 3.3	7.6 ± 2.7	4.3 ± 2.8
	MBS	2.4 ± 2.0	2.6 ± 2.0	2.2 ± 1.9	3.0 ± 2.3
	Mirada	2.3 ± 2.0	2.4 ± 1.9	2.2 ± 2.0	2.8 ± 2.0
Sub #3	VEL	3.2 ± 3.2	3.4 ± 2.7	4.4 ± 4.1	4.8 ± 4.3
	MIM	4.4 ± 3.6	4.0 ± 3.0	5.5 ± 4.7	5.0 ± 4.4
	DFD	4.7 ± 2.4	5.7 ± 3.5	6.5 ± 3.6	4.8 ± 3.4
	ILK	3.1 ± 2.7	7.1 ± 4.9	7.5 ± 5.0	4.5 ± 4.5
	IOF	2.9 ± 2.1	6.0 ± 3.9	6.8 ± 3.5	4.4 ± 3.1
	MBS	3.6 ± 3.6	3.5 ± 2.7	5.8 ± 4.6	4.4 ± 4.1
	Mirada	3.5 ± 3.4	3.1 ± 2.6	5.2 ± 4.3	4.4 ± 4.2

Table 3 demonstrates the mean TRE of seven DIR methods in three healthy subjects for the upper lung and the lower lung, respectively. The TRE is calculated in three directions ML,AP,SI ,as well as the 3D magnitude.

From Table 3, large TRE (13.3 - 23.7 mm) was found in Subject 1. For Subjects 2 and 3, TRE was 2.5 - 6.4 mm and 4.4 - 5.7 mm respectively, which closely matched the reported accuracy of DIR methods (typically 1-2 voxel sizes). The mean (max) TRE for seven DIR methods were: 7.3 (42.2) mm [VEL], 7.7 (48.2) mm [MIM], 11.2 (65.0) mm [DFD], 10.0 (79.2) mm [ILK], 10.8 (67.7) mm [IOF], 7.8 (68.3) mm [MGB] and 6.8 (74.8)

mm [Mirada]. TRE in ML direction ranged 2.3-6.1mm, while TRE in AP and SI directions ranged 2.4-21.2 mm and 2.6-28mm, respectively.

Table 3 Mean TRE (in mm) and 3D Magnitude of the seven DIR methods in three healthy subjects for the upper and lower lungs.

		ML		AP		SI		Magnitude	
		Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower
Sub #1	VEL	3.8	4.2	8.1	12	4.6	14.6	9.7	19
	MIM	4.2	4.8	7.5	9.3	5.4	16	9.1	18.7
	DFD	3.9	4.7	13.1	18.9	7.6	28	9.9	33.9
	ILK	3.5	5.3	15.1	21.2	9.2	27	9.6	33.8
	IOF	3.6	5.2	14.2	20.8	7.5	26.8	10.6	35.4
	MBS	4.8	6.1	8.7	14.9	8	18.9	10.7	20.7
	Mirada	5.7	7.5	9.8	19.6	9.2	25.6	11.9	29.7
Sub #2	VEL	2.1	2.7	2.2	2.1	1.6	2.6	2.5	2.5
	MIM	3.2	3.1	3.1	2.9	3.4	3	3.1	3.5
	DFD	4.8	4.7	3.8	4.2	9.2	6.3	6.9	5.9
	ILK	1.9	2.2	6	5.5	8.1	4.9	3.1	2.9
	IOF	3.5	3	4.4	4.6	9	6.1	4.8	3.8
	MBS	2.2	2.6	2.5	2.8	1.9	2.6	2.6	3.4
	Mirada	2.2	2.5	2.3	2.5	1.9	2.6	2.6	3
Sub #3	VEL	2.6	3.9	2.6	4.4	3.2	5.9	3	7.1
	MIM	3.8	5.2	3	5.2	4.6	6.7	4.1	7.6
	DFD	5.2	4.1	5.3	6.2	7	6	5.2	4.3
	ILK	2.3	4	6.9	7.3	7	8.2	2.5	7.1
	IOF	2.9	3.1	5.8	6.3	7.3	6.1	4.1	4.7
	MBS	3.2	4.2	3.1	4	5.1	6.6	3	6.2
	Mirada	2.9	4.3	2.4	4	4.2	6.4	2.7	6.5

3.1.2 TRE Frequency Distribution

A 3D error frequency distribution was generated as shown in Figure 6 to Figure 8 for each healthy subject. In the figure, 3D error for every DIR method was demonstrated as histogram map within error bins.

Across three subjects, the DFD and IOF methods distributed 3D error relatively evenly throughout all tagging grids. Other methods showed similar error distribution. For subject 1, larger than 50% error data falls into the category of 15 mm to 20 mm. For subject 2 and 3, about 80% error data falls into the category of 2.5 mm to 5.0 mm.

The reason for choosing bin as 2.5 mm for display is that this number matches the resolution for high-resolution MR images.

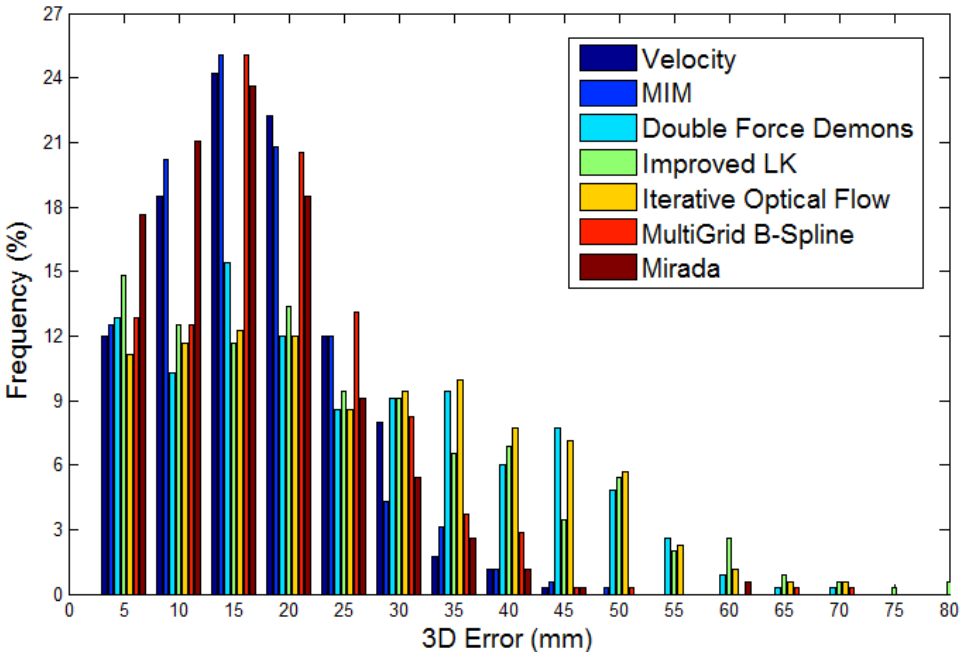

Figure 6 Frequency distribution of 3D Error for seven DIR methods in subject 1.

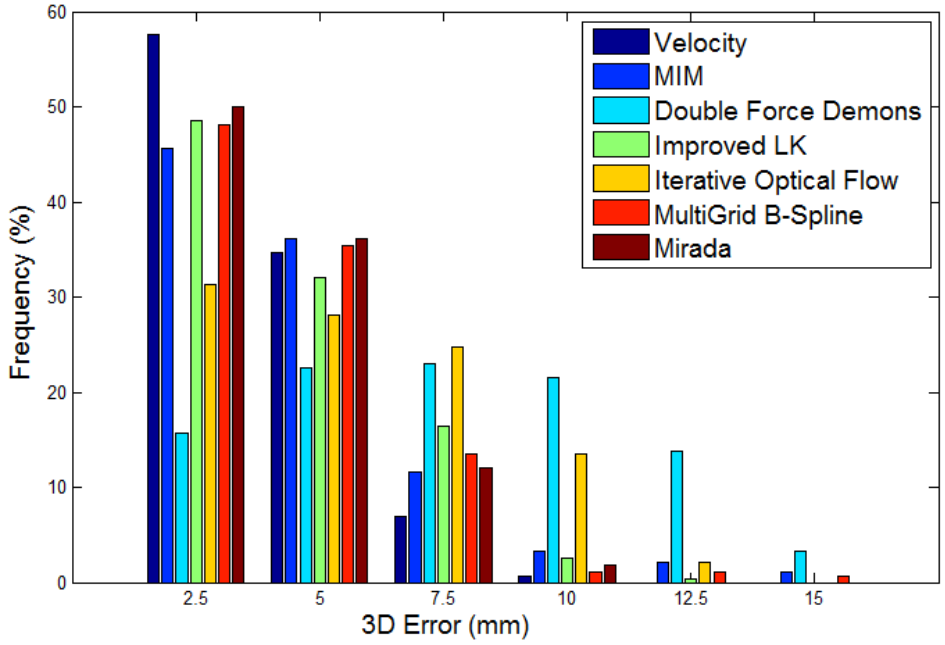

Figure 7 Frequency distribution of 3D Error for seven DIR methods in subject 2.

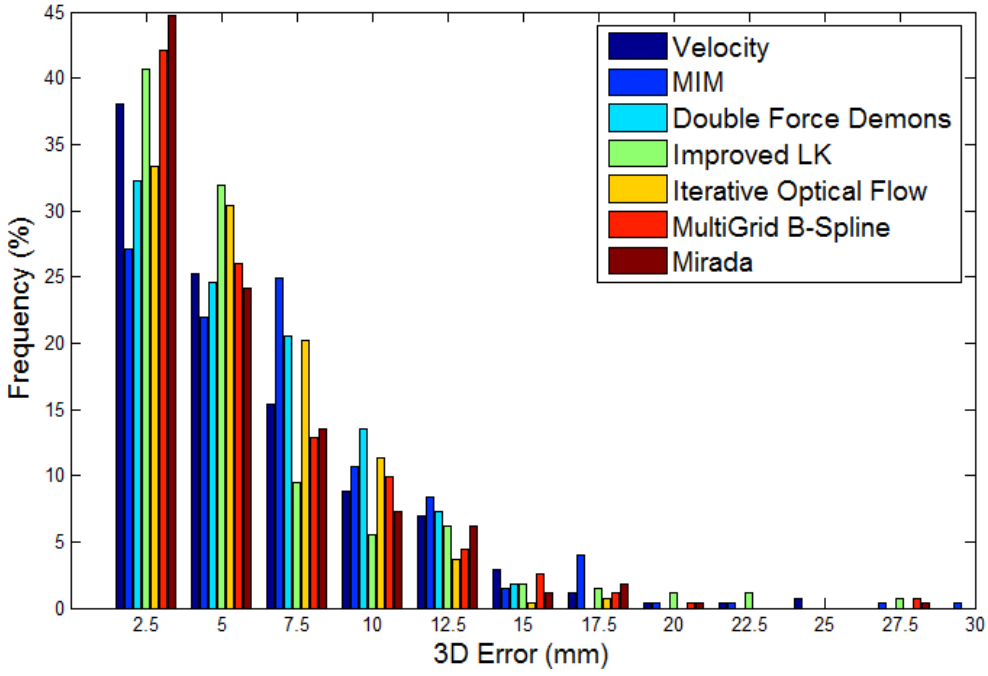

Figure 8 Frequency distribution of 3D Error for seven DIR methods in subject 3.

3.2 DVF Magnitude Comparison-Distribution and Correlation

We evaluated the measured DVF and the DIR-derived DVF in this section. Firstly, a DVF magnitude distribution maps were created for each subject separately. Secondly, a correlation study between the measured DVF and the DIR-derived DVF was performed.

3.2.1 DVF Magnitude Distribution

For DVF magnitude distribution, the DVF in 3D was sorted into different groups associated with its magnitude. The bin for the magnitude is 2mm for seven DIR methods in three healthy subjects. Figure 9 illustrates the number of tagging grids that demonstrated a 3D displacement within the limit of each bin.

We can clearly see the magnitude distribution for DFD and IOF algorithms differs from any other methods where the magnitude mostly concentrates on 10-15 mm. For Subject 1 who had large lung motion (45.6 mm), 80% of dDVFs ranged 7-30mm, while 80% tDVF ranged 15-60 mm. For Subjects 2 and 3 who had relatively small lung motions (18.1 mm, 22.2 mm, respectively), similar distributions were found between rDVF and tDVF.

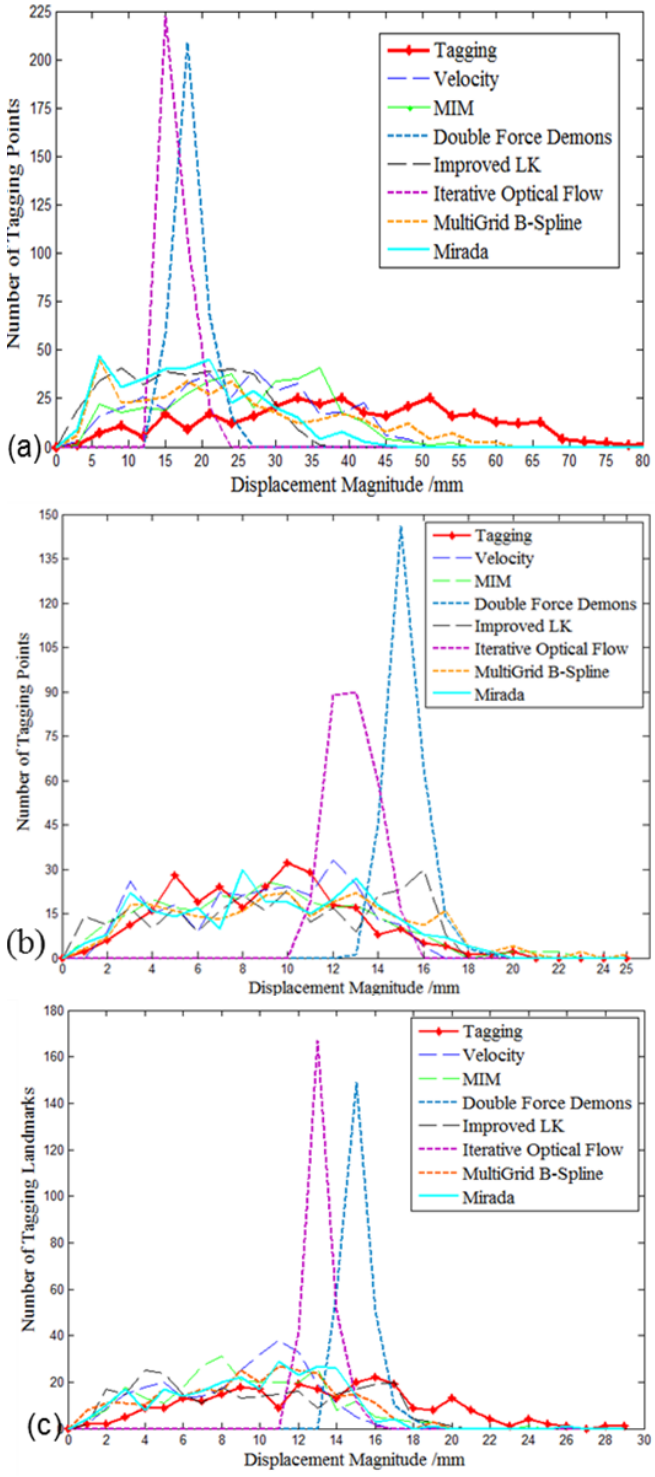

Figure 9 The total lung DVF's magnitude distributions of the tagging method and 7 DIR methods for (a) Subject 1, (b) Subject 2, and (c) Subject 3.

Table 4 summarizes correlation coefficient value for the measured 3D displacement distribution and seven DIR methods. This correlation coefficient indicates the similarity of distribution between DIR-derived displacement and measured displacement.

Table 4 Summary of correlation coefficient of 3D displacement distribution between seven DIR methods and measured 3D displacement in three subjects.

	Subject 1	Subject 2	Subject 3
VEL	0.36	0.87	0.52
MIM	0.42	0.95	0.64
DFD	0.03	0.10	0.42
ILK	0.05	0.67	0.69
IOF	0.06	0.24	0.44
MGB	0.17	0.85	0.73
Mirada	0.43	0.82	0.67

A higher correlation coefficient (0.10-0.95) was found in Subjects 2 and 3, while the correlation coefficient for Subject 1 ranges from 0.03 to 0.43. For each DIR method, VEL, MIM and Mirada maintained a relatively higher correlation coefficient (0.36-0.95) for all three subjects.

3.2.2 Correlation between Measured DVF and DIR-derived DVF

When observing the relationship between measured DVF and DIR-derived DVFs, a mild linear trend can be found as illustrated in Table 5 is just for illustration purpose, so only the plot of correlation for subject 1 was shown here. A thorough correlation coefficient value for seven DIR methods in 3 subjects can be found in Table 5. Interestingly, although all DIR methods showed a relatively low correlation coefficient (0.30-0.43) in subject 1 between DVF distribution s, all DIR methods demonstrated a relatively high correlation coefficient (0.02-0.93) for direct DVF comparison, especially for VEL, MIM and Mirada.

Table 5 Summary of correlation coefficient between measured DVF and DIR-derived DVF in three subjects.

	Subject 1	Subject 2	Subject 3
VEL	0.93	0.69	0.46
MIM	0.91	0.50	0.22
DFD	0.02	0.22	0.22
ILK	0.26	0.66	0.46
IOF	0.18	0.57	0.39
MGB	0.81	0.68	0.50
Mirada	0.89	0.67	0.49

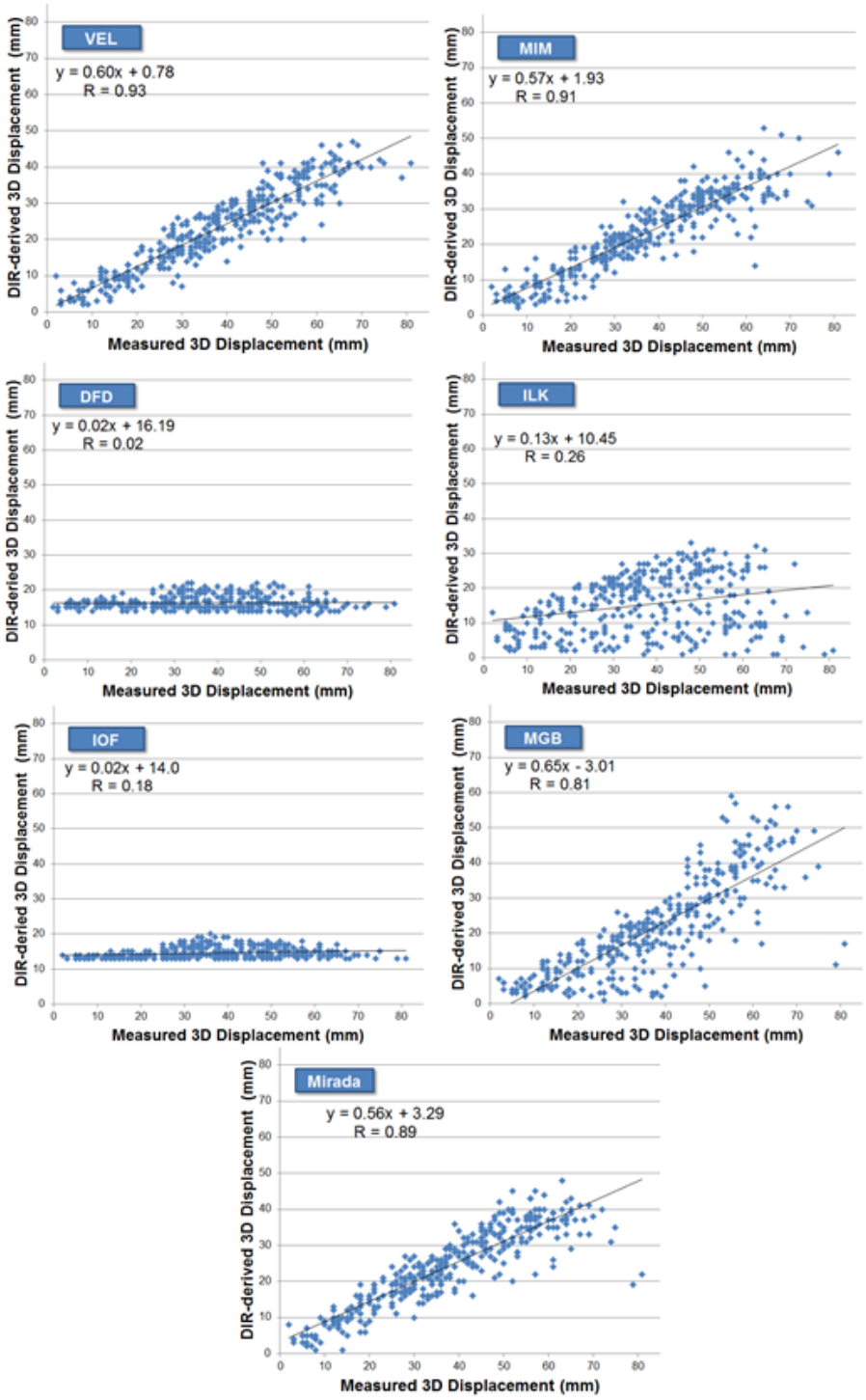

Figure 10 Illustrations for correlation study of measured 3D displacement and DIR-derived 3D displacement for tagging grids in subject 1.

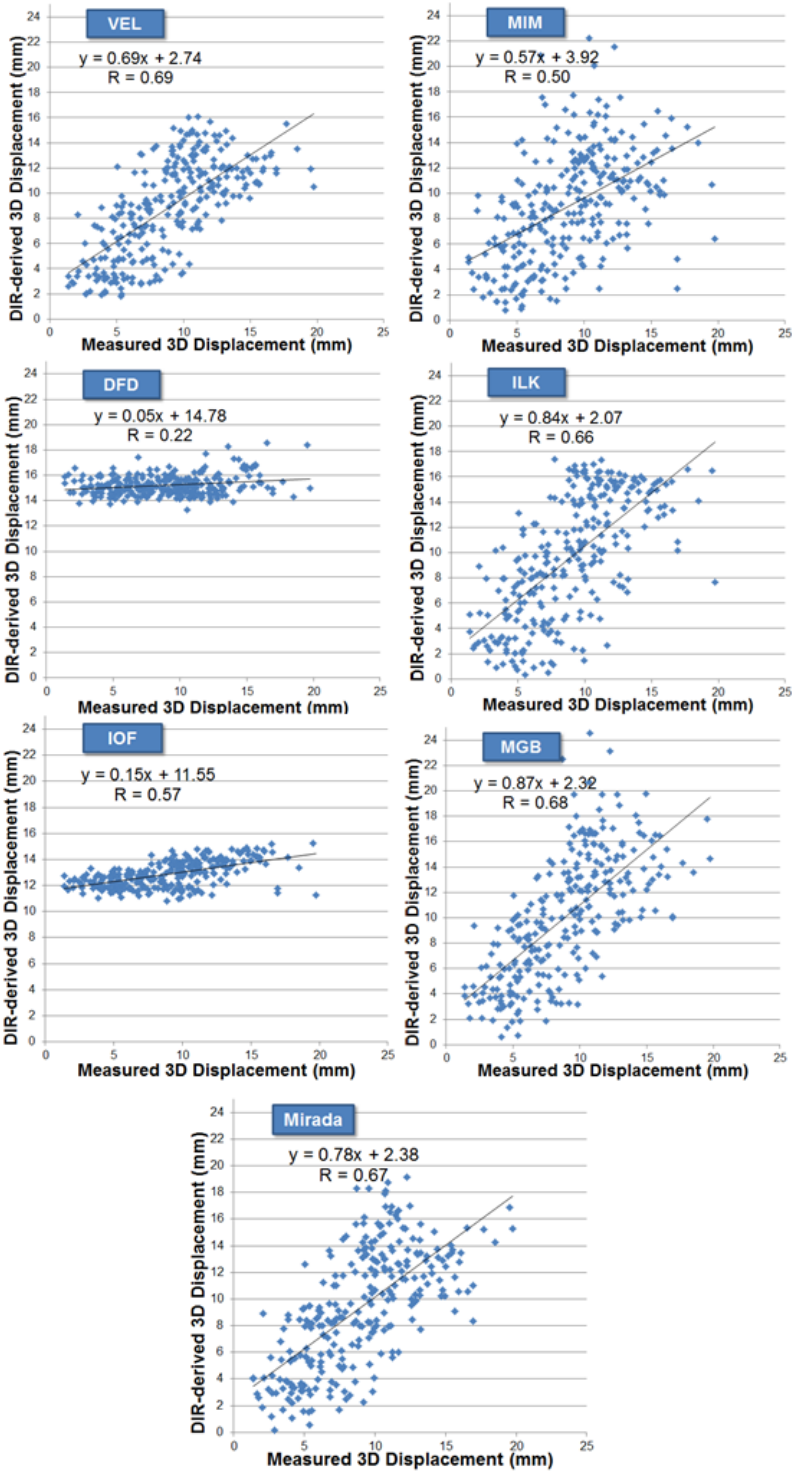

Figure 11 Illustrations for correlation study of measured 3D displacement and DIR-derived 3D displacement for tagging grids in subject 2.

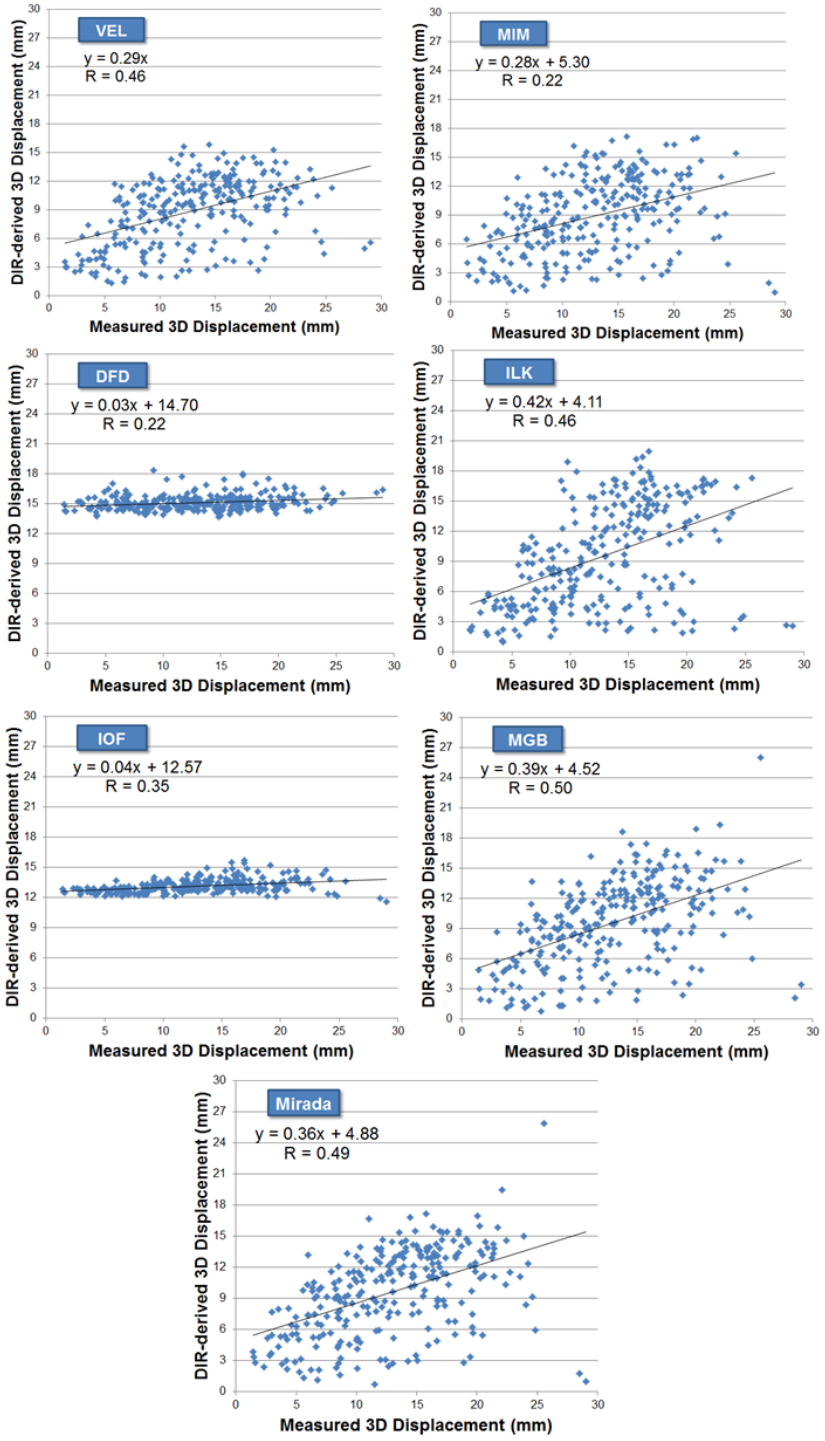

Figure 12 Illustrations for correlation study of measured 3D displacement and DIR-derived 3D displacement for tagging grids in subject 3.

3.3 Spatial Error Distribution

TRE distribution in space was demonstrated in Figure 13 to Figure 15 in TRE color maps.

For each subject, the TRE in ML direction was illustrated in sagittal view and the TRE in SI direction was illustrated in coronal view.

In SI direction, it was observed that the lower lung in general demonstrates larger TRE than the upper lung, especially in Subject 1. In the ML direction, no single conclusion can be made from this visualization. But in subject 2 for MGB and Mirada TRE color map, higher TRE can be observed for the lung fissure.

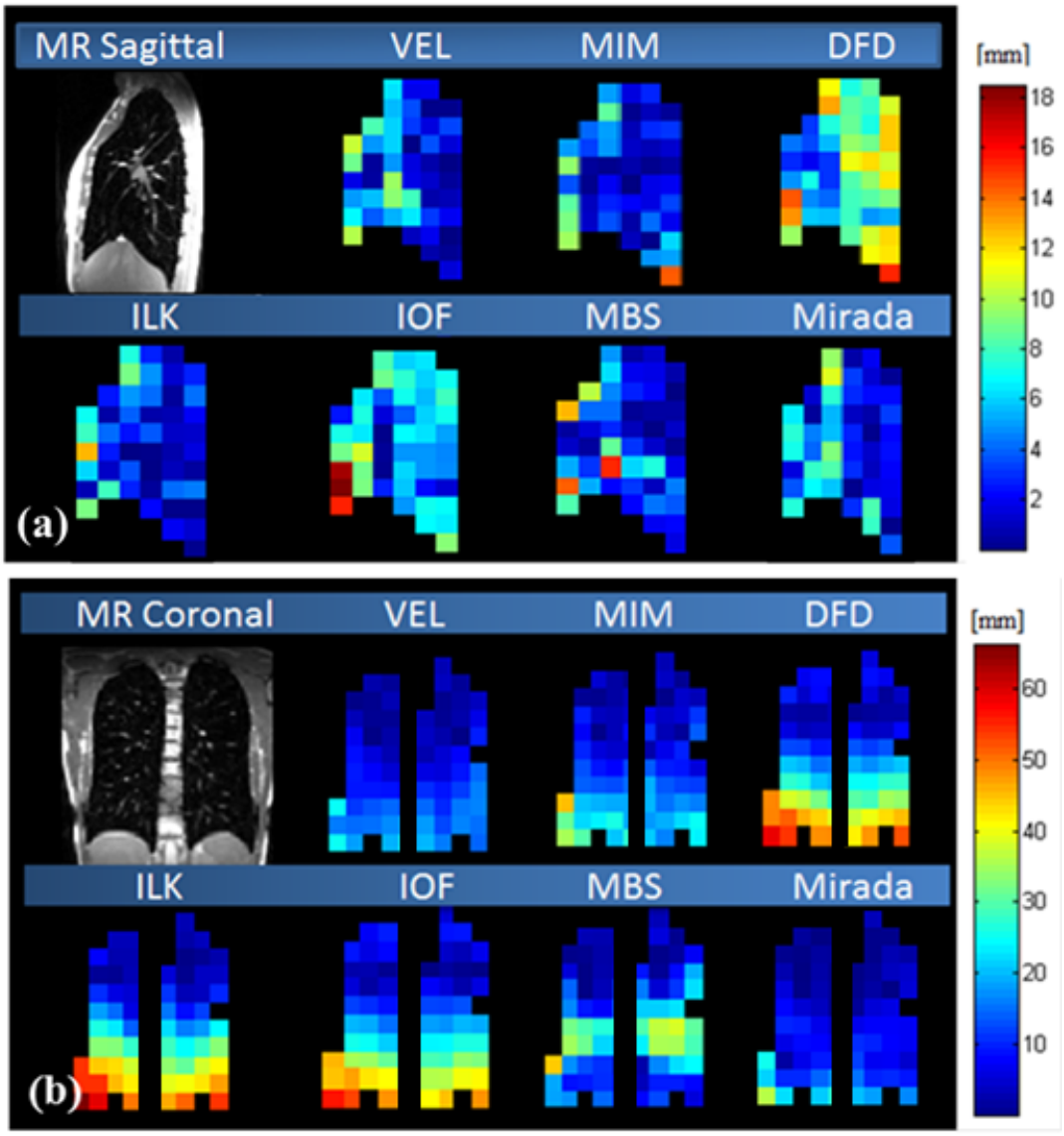

Figure 13 Examples of TRE maps in the ML direction in Sagittal View (a) and SI direction in Coronal View (b) for the 7 DIR methods in Subject 1.

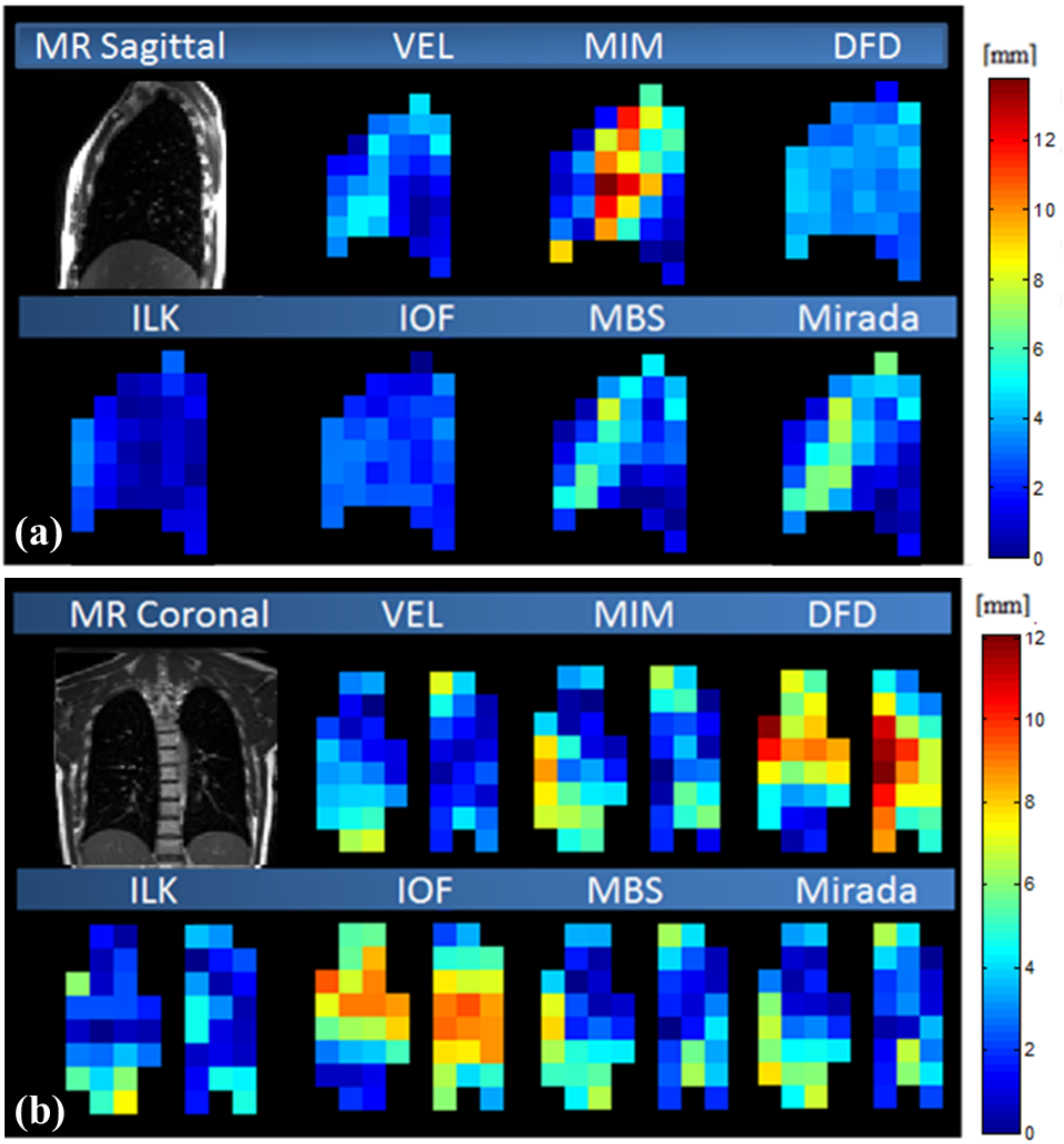

Figure 14 Examples of TRE maps in the ML direction in Sagittal View (a) and SI direction in Coronal View (b) for the 7 DIR methods in Subject 2.

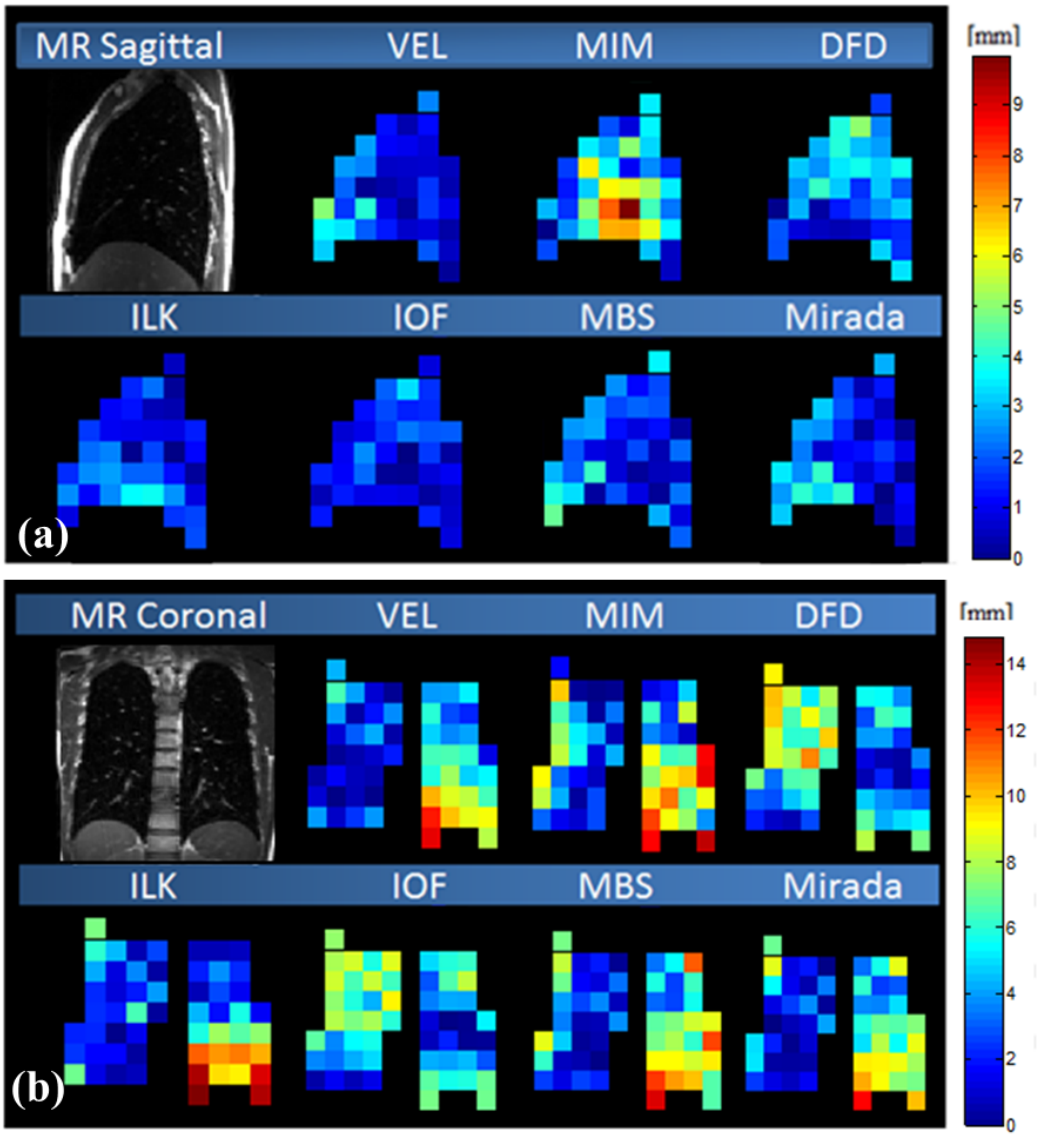

Figure 15 Examples of TRE maps in the ML direction in Sagittal View (a) and SI direction in Coronal View (b) for the 7 DIR methods in Subject 3.

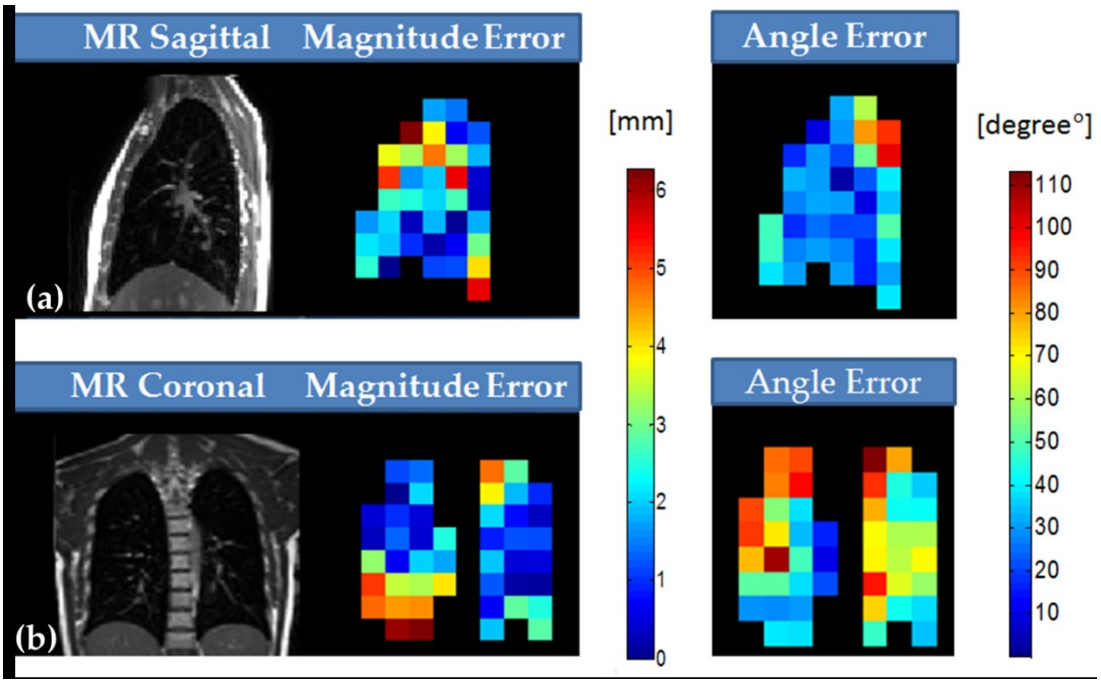

Figure 16 Examples of TRE maps of magnitude error and angle error in Sagittal View (a) and Coronal View (b) for VEL method in Subject 2.

3.4 A Correlation Study on TRE and Lung Motion

Figure 17 illustrates the correlation between the measured 3D motion and the 3D TRE for the seven DIR methods. In general the linear relationship between the two was observed (r^2=0.61- r^2=0.82).

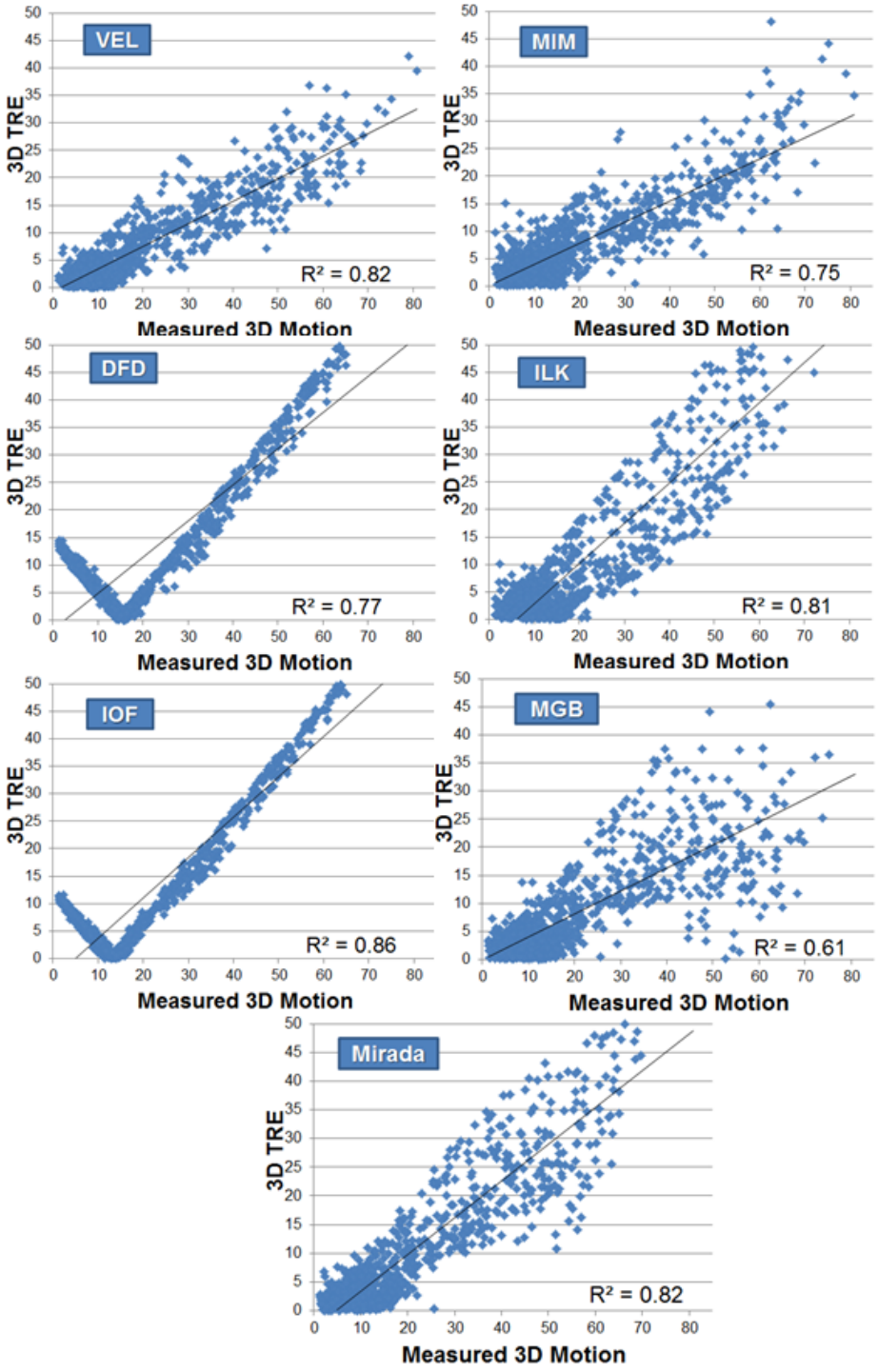

Figure 17 Correlation between the measured 3D displacements to 3D TRE for seven DIR methods.

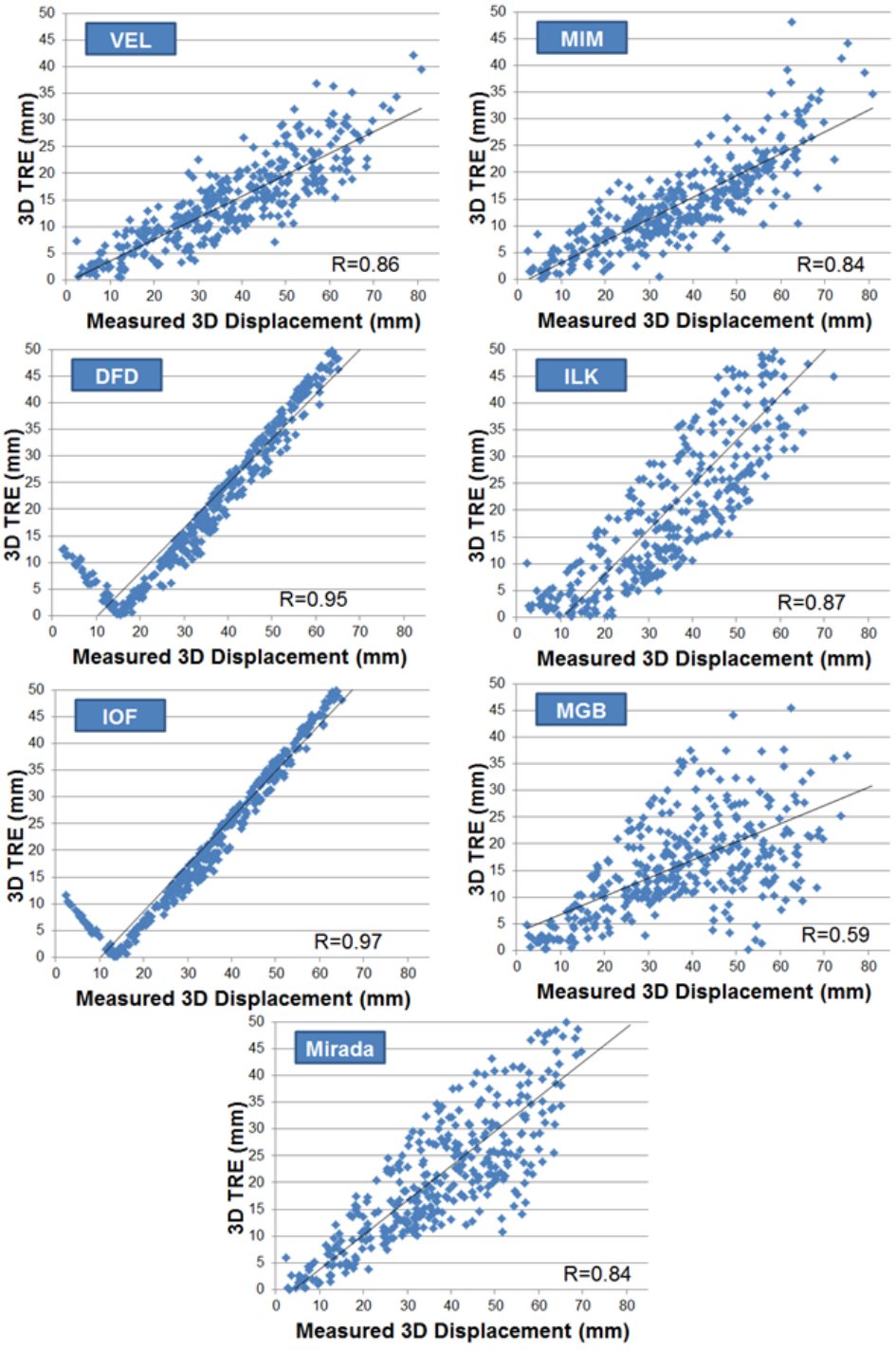

Figure 18 Correlation between the measured 3D displacements to 3D TRE for seven DIR methods for Subject 1.

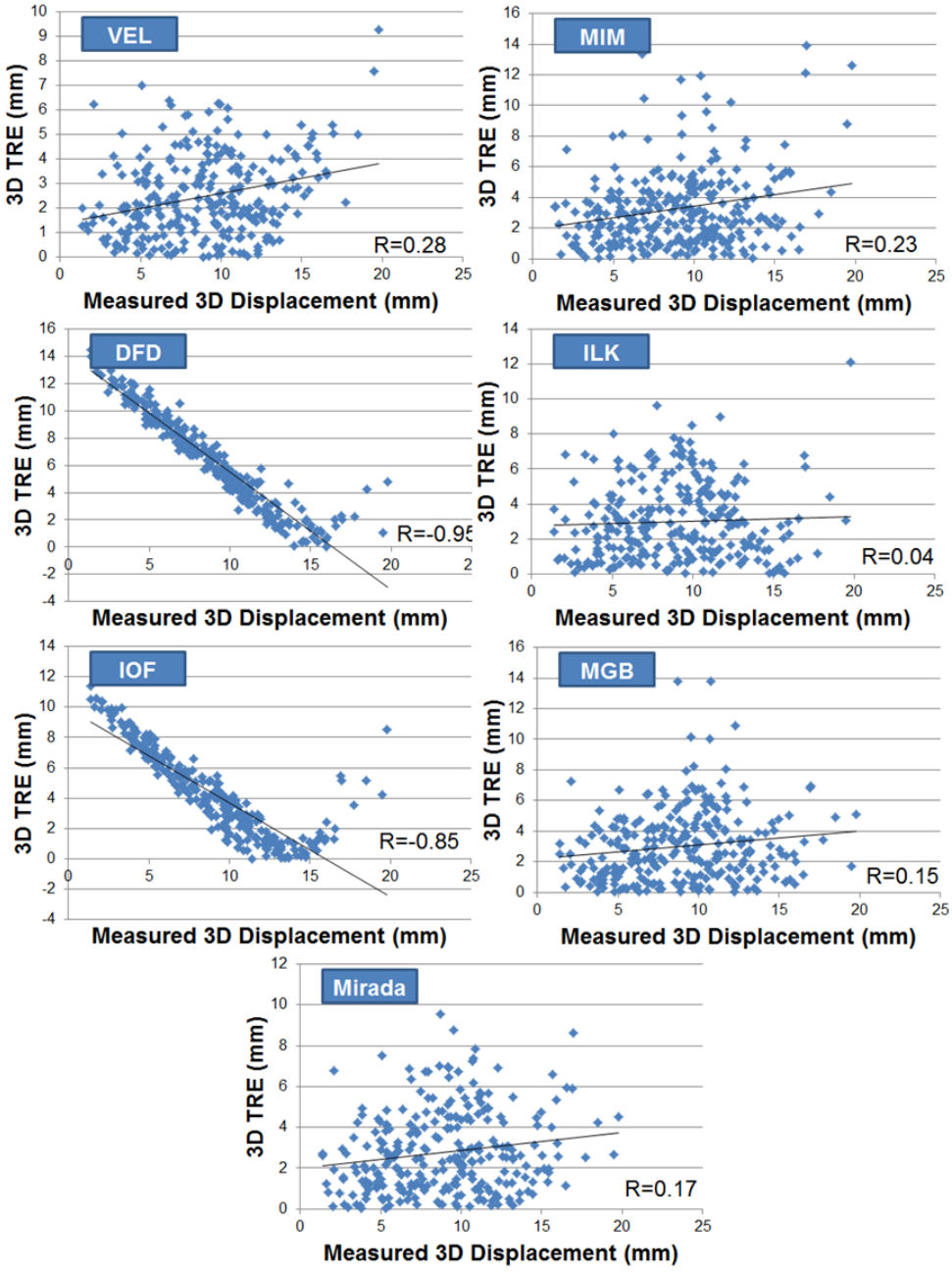

Figure 19 Correlation between the measured 3D displacements to 3D TRE for seven DIR methods for Subject 2.

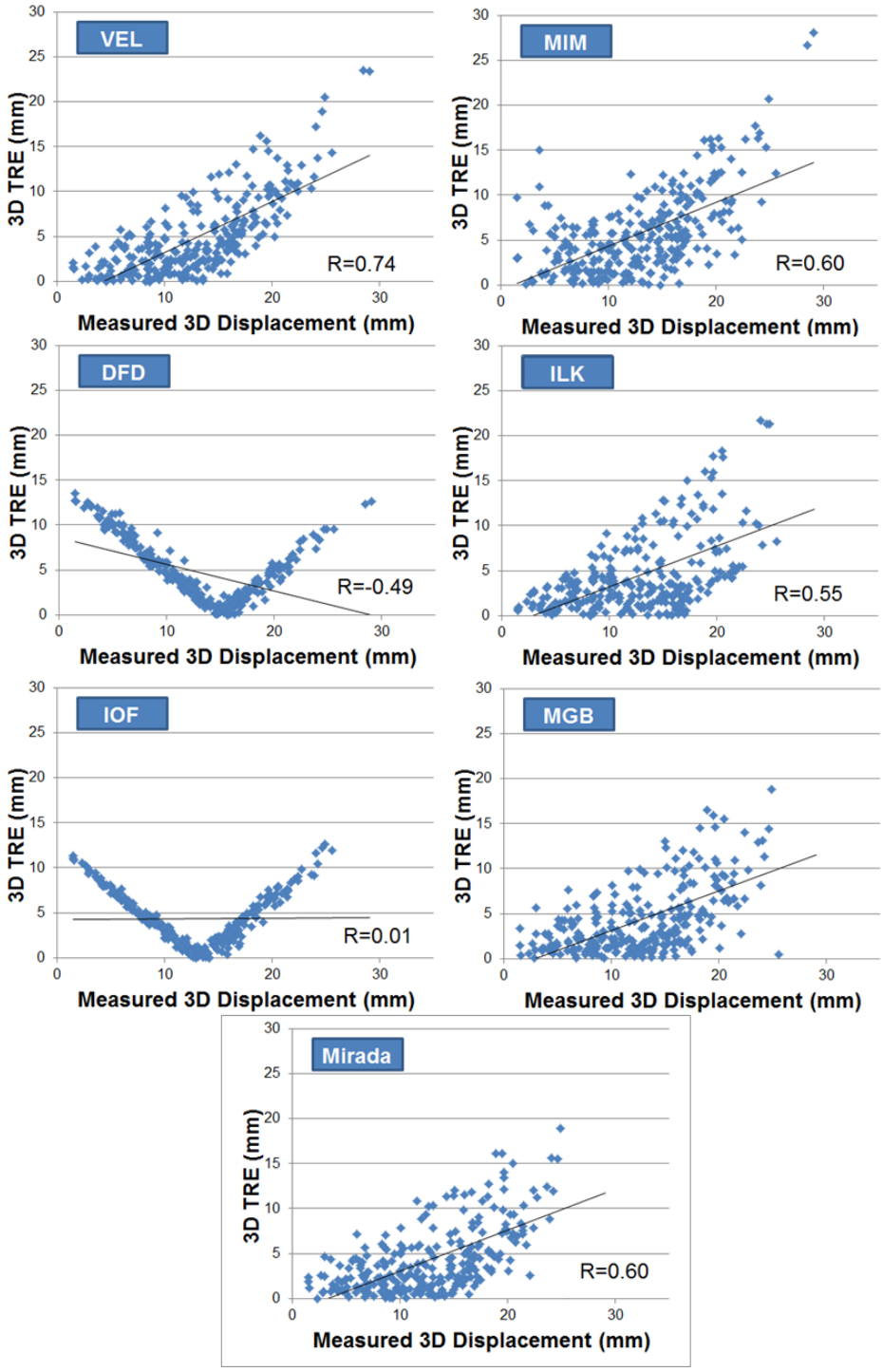

Figure 20 Correlation between the measured 3D displacements to 3D TRE for seven DIR methods for Subject 3.

It should be noted here the check mark shape DFD and IOF demonstrated instead of a scattered linear distribution. As we have observed from Figure 9, these two methods showed a uniform DVF distribution over the tagging grids, which means the dDVF is a constant number. TRE was defined as |tDVF-dDVF|. So if the dDVF basically remains constant, the relationship between TRE and measured displacement which is tDVF will demonstrate a check mark shape. The point value where the line intersects with the x-axis is the constant dDVF value which is about 15mm for both methods.

3.5 Uncertainties in Evaluation Study Design

Though carefully monitored, variation and errors were introduced in the evaluation study design. These errors can be classified into two categories, tagging image diffusion error, and human error.

Tagging image diffusion error: When the subject inhales the hyperpolarized gas, the gas would diffuse during the breathing cycle. The average diffusion for He and Xe is 0.2 cm^2/s and 0.04 cm^2/s. [19] The diffusion could potentially influence the tagging point size. However, the diffusion is homogeneous in the lung. And we recorded the center of the tagging point. As long as the gas diffused at the same speed in all directions, the center location would not change. This error is negligible.

Human error: The tagging grid position is manually recorded. Thus human error in interpreting central point was introduced. The error was defined by calculating

variation when identifying center point position multiple times. This error is approximately 0.7mm in all three directions.

4. Discussion

4.1 Interpretation of Results

In this study we quantitatively evaluated the accuracy of deformable image registration (DIR) in lung motion estimation using HP gas tagging MRI as references. Significant differences were observed in lung motion estimation between the HP gas tagging MRI method and the DIR methods. Variations in the performance of DIR methods were also observed.

DVF (in three directions and magnitude) is the major variable we look into and compare. Direct DVF comparison which involves TRE was performed first. Next distribution maps were drawn to gain insight into the statistics. Large TRE (13.3-23.7 mm) was found in Subject 1. For Subjects 2 and 3, TRE was 2.5-6.4 mm and 4.4-5.7 mm respectively, which closely matched the reported accuracy of DIR methods (typically 1-2 voxel sizes).

In this study we investigated the correlation of measured 3D displacement, which indicates lung motion, and the registration error. A strong correlation was observed (r=0.78-0.91). This meets our expectation that the lung motion magnitude has an effect on the DIR accuracy.

The rationale for investigating measured 3D displacement as influence factor is that the displacement is the only data sufficient enough to provide statistical results

based on three healthy subjects. For each subject, the number of tagging grid ranges from 274 to 351. So the data scale of measured 3D displacement is in the order of 1000.

However, the displacement motion is only one characteristic of lung deformation during breathing. There are other factors that would represent the lung deformation including volume changing (ventilation map), density changing or breathing irregularity. These factors can be represented by variable that is subject-independent. In this study, only three subjects were involved, so no significant correlation was observed. Here we summarized some of these factors we calculated for three subjects just for demonstration.

Table 6 summarized the mean TRE, mean DVF, DVF standard deviation, DVF coefficient of variation in 3D and SI direction respectively. The coefficient of variation is defined as

$$C_v = \frac{\sigma}{\mu}$$

The coefficient of variation will represent the variability of entire lung motion, which reveals the breathing variability over the lung.

Table 6 The mean TRE (mm) in 3D and in SI direction. Mean DVF μ_{DVF} (mm) in 3D and in SI direction. DVF standard deviation σ_{DVF} (mm) in 3D and in SI direction. Coefficient of variation for DVF C_{v-DVF} in 3D and in SI direction.

	3D				SI			
	μ_{TRE}	μ_{DVF}	σ_{DVF}	C_{v-DVF}	μ_{TRE-SI}	μ_{DVF-SI}	σ_{DVF-SI}	$C_{v-DVF-SI}$
Sub # 1	19.22	37.26	17.05	0.46	15.30	24.16	19.81	0.82
Sub # 2	3.61	8.92	3.79	0.43	5.23	2.27	3.37	1.48
Sub # 3	4.71	12.83	5.55	0.43	6.04	7.12	2.78	0.39

There is no strong conclusion on the relationship between the mean registration error and the coefficient of variation based on limited data sets we had. But from the trend we demonstrated above, a mild conclusion that the TRE has decent relationship with DVF magnitude and variance can be made. The variability over the lung doesn't influence the registration error significantly. The main concern that would introduce large registration error will still be breathing amplitude.

4.2 Limitations of the Study

The limitations of this study are mainly from two parts- the technique limitation and the study design limitation.

The hyperpolarized tagging technique we used here is only available to image the subject in two phases during a breathing cycle due to the longer scanning time. As we have stated before, a strong correlation between the breathing motion and registration error was observed. If we only image the subject in EOI and EOE phase, the

breathing motion would be large compared to 10 phases imaging within one breathing cycle, thus introducing larger registration error. In addition, this method would be potentially useful in 4D radiation therapy DIR algorithm validation and dose calculation. The real time imaging technique or the technique that acquire images more than two phases within one breathing cycle will evaluate the real registration error propagation and distribution more practically.

This study design involves 3 sets of image data. Tagging MR and low-resolution proton MR acquired at the same time, and high-resolution proton MR acquired at two different breath holds. All the registration is based on the assumption that the subject's position and breathing pattern is exactly the same for all three image data acquisition. We limited the breathing motion by providing same amount of air for inhalation and exhalation during image acquisition scanning. However, due to the technical limitation and uncertainty in breathing pattern, a small variation is observed during scanning. A thorough study into how this variation between the high-resolution proton MR and low-resolution proton MR could influence the registration error need to be evaluated.

5. Conclusion

A methodology for evaluating lung DIR using tagging MR has been developed.

These preliminary results demonstrated that grid tagging HP gas MRI is a unique, effective method for quantitatively evaluating deformation image registration of the lungs. Tagging MRI provides accurate motion information for hundreds of internal physiological landmarks in the lungs that can be used to quantitatively evaluate deformation image registrations of the lungs.

Differences were observed in lung motion estimation between the HP gas tagging MRI method and the DIR methods depending on the subject breathing pattern. A linear correlation was found between the breathing motion and the registration error. The difference was found greater when the lung motion is larger. Variations in lung motion estimation were observed among different DIR methods.

References

1. Varadhan R, Karangelis G, Krishnan K, Hui S.A framework for deformable image registration validation in radiotherapy clinical applications. J Appl Clin Med Phys. 2013 Jan 2;14(1):4066.
2. Hardcastle N et al, Accuracy of deformable image registration for contour propagation in adaptive lung radiotherapy. Radiation Oncology 2013 8:243.
3. Hardcastle N, Tomé WA, Cannon DM, Brouwer CL, et al. A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy. Radiat Oncol. 2012 Jun 15;7:90.
4. Kadoya N, Fujita Y, Katsuta Y, Dobashi S, et al. Evaluation of various deformable image registration algorithms for thoracic images. J Radiat Res. 2014 Jan 1;55(1):175-82
5. Zhou L, Zhou L, Zhang S, Zhen X, Yu H, Zhang G, Wang R. Validation of an improved 'diffeomorphic demons' algorithm for deformable image registration in image-guided radiation therapy. Biomed Mater Eng. 2014;24(1):373-82.
6. Brock KK.Results of a Multi-Institution Deformable Registration Accuracy Study (MIDRAS). Int J Radiat Oncol Biol Phys. 2010 Feb 1;76(2):583-96
7. Castadot P, Lee JA, Parraga A, Geets X, Macq B, Grégoire V. Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors. Radiother Oncol. 2008 Oct;89(1):1-12.
8. Kirby N, Chuang C, Ueda U, Pouliot J. The need for application-based adaptation of deformable image registration. Med Phys. 2013 Jan;40(1):011702.
9. Latifi K, Zhang G, Stawicki M, van Elmpt W, Dekker A, Forster K.Validation of three deformable image registration algorithms for the thorax. J Appl Clin Med Phys. 2013 Jan 7;14(1):3834.
10. Kashani R, Hub M, Balter JM, et al.Objective assessment of deformable image registration in radiotherapy: A multi-institution study. Med Phys. 2008 Dec;35(12):5944-53.
11. Cai J, Altes TA, Miller GW, et al.MR Grid-Tagging Using Hyperpolarized Helium-3 for Regional Quantitative Assessment of Pulmonary Biomechanics and Ventilation. Magn Reson Med. 2007 Aug;58(2):373-80.

12. Cai J, Miller GW, Altes TA, et al. Direct measurement of lung motion using hyperpolarized helium-3 MR tagging.Int J Radiat Oncol Biol Phys. 2007 Jul 1;68(3):650-3.

13. Cai J, Sheng K, Miller GW, et al. Dynamic MRI of grid-tagged hyperpolarized helium-3 for the assessment of lung motion during breathing. Int J Radiat Oncol Biol Phys. 2009 Sep 1;75(1):276-84.

14. S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P.W. Pluim, "elastix: a toolbox for intensity based medical image registration," IEEE Transactions on Medical Imaging, vol. 29, no. 1, pp. 196 - 205, January 2010.

15. D.P. Shamonin, E.E. Bron, B.P.F. Lelieveldt, M. Smits, S. Klein and M. Staring, "Fast Parallel Image Registration on CPU and GPU for Diagnostic Classification of Alzheimer's Disease", Frontiers in Neuroinformatics, vol. 7, no. 50, pp. 1-15, January 2014

16. Yang D, Brame S, El Naqa I, Aditya A, Wu Y, Goddu SM, Mutic S, Deasy JO, Low DA.Technical note: DIRART--A software suite for deformable image registration and adaptive radiotherapy research. Med Phys. 2011 Jan;38(1):67-77.

17. Sotiras A, Davatzikos C, Paragios N, IEEE Trans Med Imaging. Jul 2013; 32(7): 1153–1190.

18. Yang D, Naqa I.DIRART (Deformable Image Registration and Adaptive Radiotherapy) Software Suite User Instruction Manual Version 0.1 2009

19. Mugler J, Altes T, Hyperpolarized 129Xe MRI of the Human Lung. J Magn Reson Imaging. Feb 2013; 37(2): 313–331.