

## Using patient-specific phantoms to evaluate deformable image registration algorithms for adaptive radiation therapy

Nick Stanley,<sup>1</sup> Carri Glide-Hurst,<sup>1</sup> Jinkoo Kim,<sup>1</sup> Jeffrey Adams,<sup>2</sup> Shunshan Li,<sup>1</sup> Ning Wen,<sup>1</sup> Indrin J Chetty,<sup>1</sup> Hualiang Zhong<sup>1a</sup>  
*Department of Radiation Oncology,<sup>1</sup> Henry Ford Health System, Detroit, MI, USA;*  
*Department of Radiation Oncology,<sup>2</sup> Wayne State University, Detroit, MI, USA*  
*hzhong1@hfhs.org*

Received 11 January, 2013; accepted 14 June, 2013

The quality of adaptive treatment planning depends on the accuracy of its underlying deformable image registration (DIR). The purpose of this study is to evaluate the performance of two DIR algorithms, B-spline-based deformable multipass (DMP) and deformable demons (Demons), implemented in a commercial software package. Evaluations were conducted using both computational and physical deformable phantoms. Based on a finite element method (FEM), a total of 11 computational models were developed from a set of CT images acquired from four lung and one prostate cancer patients. FEM generated displacement vector fields (DVF) were used to construct the lung and prostate image phantoms. Based on a fast-Fourier transform technique, image noise power spectrum was incorporated into the prostate image phantoms to create simulated CBCT images. The FEM-DVF served as a gold standard for verification of the two registration algorithms performed on these phantoms. The registration algorithms were also evaluated at the homologous points quantified in the CT images of a physical lung phantom. The results indicated that the mean errors of the DMP algorithm were in the range of 1.0–3.1 mm for the computational phantoms and 1.9 mm for the physical lung phantom. For the computational prostate phantoms, the corresponding mean error was 1.0–1.9 mm in the prostate, 1.9–2.4 mm in the rectum, and 1.8–2.1 mm over the entire patient body. Sinusoidal errors induced by B-spline interpolations were observed in all the displacement profiles of the DMP registrations. Regions of large displacements were observed to have more registration errors. Patient-specific FEM models have been developed to evaluate the DIR algorithms implemented in the commercial software package. It has been found that the accuracy of these algorithms is patient-dependent and related to various factors including tissue deformation magnitudes and image intensity gradients across the regions of interest. This may suggest that DIR algorithms need to be verified for each registration instance when implementing adaptive radiation therapy.

PACS numbers: 87.10.Kn, 87.55.km, 87.55.Qr, 87.57.nj

Key words: deformable image registration, validation, finite element modeling, deformable phantom

Conflict of Interest statement: The authors declare there are no conflicts of interest.

### I. INTRODUCTION

Anatomy revealed in a planning CT image may change during the course of radiation treatment due to factors such as patient breathing, setup errors, or patient weight loss.<sup>(1,2)</sup> Anatomical

<sup>a</sup> Corresponding author: Hualiang Zhong, Department of Radiation Oncology, Henry Ford Health System, 2799 W Grand Blvd., Detroit, MI 48202, USA; phone: 1 (313) 876 9180; fax: 1 (313) 916 3264; email: hzhong1@hfhs.org

changes may compromise the accuracy of dose calculation for each organ. As a result, the treatment plan developed may not provide necessary target coverage and organ-at-risk sparing. Adaptive radiotherapy (ART) aims to minimize the dosimetric impact of anatomical changes by reoptimizing the original treatment plan if its quality degrades.<sup>(3)</sup> A key step in the implementation of ART is to match each point on daily CT images to their correspondent points in the planning image. This process can be accomplished with deformable image registration (DIR) techniques.<sup>(4)</sup> DIR is to derive a transformation map by maximizing the intensity similarity between the two images being registered. Depending on the registration techniques used, the transformation map can be represented by different mathematical models such as affine transform,<sup>(5)</sup> thin-plate spline,<sup>(6)</sup> or B-spline basis,<sup>(7)</sup> or adapted directly through optical flow-based equations.<sup>(8)</sup> Similarity metrics also can be represented in different forms including the sum of squared difference, cross-correlation, or normalized mutual information.<sup>(9,10)</sup> Like dose calculation algorithms, these DIR algorithms must be thoroughly evaluated before they are used in clinic for adaptive radiation therapy.

Visual evaluation is common practice after performing an image registration. As a preliminary test, this evaluation is convenient and especially useful for software development.<sup>(11,12)</sup> Along this direction, many efforts have been made using landmarks or contours to estimate errors in the displacement vector field (DVF)<sup>(13,14)</sup> of DIR. For example, Hardcastle et al.<sup>(15)</sup> evaluated two registration algorithms with dice scores calculated on the contours drawn by physicians, and Brock et al.<sup>(16)</sup> evaluated 4D CT registrations reported from 21 institutions by comparing the computer-predicted displacement at each bifurcation point with the displacement computed from the oncologists' annotations. Castillo et al.<sup>(17)</sup> developed an automatic method to identify and track landmark points in lung patient datasets. These studies provided quantitative evaluation results on the performances of different DIR algorithms at these distinctive landmarks or their nearby regions. However, as reported by Kashani et al.<sup>(18)</sup> and Liu et al.,<sup>(19)</sup> large registration errors can be observed in regions of uniform image intensity, and the above evaluations are limited by the number of the objects being tracked; so errors estimated by the feature-guided evaluation methods may not be representative of the registration accuracy in voxels at a distance from those landmarks.<sup>(20)</sup>

Since landmarks or contours are not always available in high-dose gradient regions, mathematical properties of the generated deformation maps may serve as an alternative metric to evaluate the quality of DIR in these regions. For example, Schreibmann et al.<sup>(21)</sup> evaluated the quality of image registrations by calculating the curl of their deformation maps. Zhong et al.<sup>(22)</sup> proposed the concept of unbalanced energy, calculated directly from DVF, to detect DIR errors. Bender and Tome<sup>(23)</sup> employed consistency metrics to evaluate the accuracy of the composed deformation maps. Klein et al.<sup>(24)</sup> used permutation and ANOVA tests to compare the relative performance of 14 nonlinear registration algorithms. These studies help evaluate the overall quality of different registrations and the derived information is valuable to clinic. However, as radiation dose was mapped by the derived DVF, subvoxel displacement errors may cause the accumulated dose over- or underestimated, especially when the dose was counted on individual particles during the dose mapping process.<sup>(25,26)</sup> Salguero et al.<sup>(27)</sup> demonstrated that registration errors greater than 1 mm can induce large dose errors in high-dose gradient regions, so the accuracy of the registration algorithm needs to be quantified at each image voxel in these regions. In addition, the performance of a registration algorithm should be evaluated for patient-specific registration scenarios that may involve different image qualities, anatomy patterns (e.g., size of homogeneous organs), and deformation magnitudes.

Computational modeling can help evaluate registration performance under various simulated scenarios.<sup>(28,29)</sup> For example, Wang et al.<sup>(30)</sup> used a set of B-spline-based mathematical phantoms to evaluate their demons algorithm, and Liu et al.<sup>(19)</sup> evaluated their DIR algorithms using the computational phantom NCAT. While the NCAT phantom has organs assigned with uniform intensity, the deformation of each organ was realistically modeled with B-spline functions. Furthermore, DIR algorithms could be verified with more realistic images acquired from

deformable physical phantoms.<sup>(19,31-33)</sup> Recently Nie et al.<sup>(34)</sup> used a set of computer-simulated deformable phantoms and a physical pelvic phantom to evaluate different DIR algorithms, and demonstrated the impact of different deformations on the performance of these algorithms. Their computational phantoms were generated by the commercial software ImSimQA (Oncology Systems Limited, Shrewsbury, UK). The global deformation of these phantoms was interpolated from the thin-plate splines which were guided by a set of manually selected control points.<sup>(35)</sup> Consequently, mass volume may not be preserved during the spline interpolation, and the resultant organ deformation may not be physically realistic.

In this study, we will first develop more realistic computational phantoms from CT images of lung and prostate patients. The images of the prostate phantoms were enhanced by the simulated CBCT noise. The reality of the phantom deformations is achieved through an in-house developed finite element modeling framework where tissue elasticity and volume change in each element were characterized by Young's modulus and Poisson ratio, and the global deformation was controlled by the conservation of tissue elasticity energy and external work. Taking these phantoms as a ground truth, we will investigate parameter settings for the DIR algorithms implemented in Velocity Advanced Imaging (VelocityAI), a commercial DIR software package, and compare the performance of these DIR algorithms for five cancer patients under different deformation scenarios. In addition, we will also develop a motor-controlled deformable physical phantom to verify these algorithms.

## II. MATERIALS AND METHODS

### A. Image registration software package

The VelocityAI software package (v2.6.2; Velocity Medical Solutions, Atlanta, GA) offers a rigid registration and three choices for 3D deformable registration algorithms: "Deformable Demons" (Demons), "Deformable Single-Pass" (DSP), and "Deformable Multi-Pass" (DMP). DMP uses DSP with preprogrammed choices for grid settings for each pass. Specifics of the grid settings are unknown to the authors as VelocityAI is a commercial, proprietary software. However, it is known that DMP uses mutual information as its similarity metric and B-splines for interpolation. Operators also have the option of limiting the registration space to a region of interest (ROI). Registration can be performed intra- and intermodality for CT, CBCT, and MRI, and its derived DVF can be exported as a binary file with three-dimensional values for each voxel.

The purpose of this study is to evaluate the DMP and Demons algorithms implemented in VelocityAI. Different from DMP, DSP allows users to choose different grid settings ranging from "coarse" to "fine". However, due to the lack of a ground truth, these settings are hardly to be used in clinic. In this study, we will first use the developed computational phantoms to evaluate different grid settings for DSP and compare their registration results with those derived from the DMP algorithm, and then we will focus on the comparison between DMP and Demons for different scenarios of organ deformation.

### B. Evaluation of image registration algorithms with lung CT images

#### *B.1 Development of computational lung phantoms*

4D CT datasets used in this study were selected from four lung cancer patients under a retrospective protocol approved by the Institutional Review Board of our institution (IRB#: 6203). For each case studied, the 3D image of the 4D CT at the end inhalation (EI) phase was considered the primary image. Computational phantoms were developed based on the primary images using a finite element modeling (FEM) system. The mathematical implementation of this system was described in our previous study.<sup>(29)</sup> With this system, tetrahedral meshes were generated and scaled to match each of the image domains. Diaphragm, spinal cord, and ribs on each side were manually segmented from the CT images. Tetrahedral nodes located in diaphragm regions

were selected as driving nodes, and those in spinal cords and side ribs were fixed as boundary constraints. Young's moduli were set to 1 MPa for ribs, 1 kPa for lung, and 10 kPa for other soft tissue. The Poisson ratio was 0.38 for lung and 0.49 for other elements.<sup>(36)</sup> With different forces assigned to the driving nodes, the displacement vectors of other anatomical structures were computed using the FEM modeling system.

For lung patient 1 (denoted Lung1), the selected nodes were assigned with the forces of 0.8, 1.6, 2.4, and 3.2 kPa, and these forces got the diaphragm moved superiorly by about 1, 2, 3, and 4 cm. The force-induced organ deformation can be visualized through the overlay of the patient's original image set and its deformed image set, as shown in Fig. 1. For lung patient 2 (denoted Lung2), the forces of 0.8, 1.2, and 2.0 kPa caused diaphragm deformation by 1.8, 2.7, and 4.2 cm, respectively. These phantoms will be used to evaluate the impact of motion magnitudes on registration accuracy.

To evaluate the impact of patient-specific anatomy on DMP registrations, CT images of two additional lung patients (Lung3 and Lung4) were included to develop more computational phantoms. 1.8 and 2.1 kPa forces were assigned to the tetrahedral nodes located in their diaphragm regions, and their lateral ribs and spinal cords were kept stationary as boundary conditions. As a result, their simulated diaphragm deformation is limited to 3 cm. Consequently, there are total four phantoms, including Lung1 and Lung2, having their diaphragms deformed about 3 cm.

The displacement vectors of the tetrahedral nodes generated by the FEM modeling system were interpolated to generate a DVF for each primary image. Secondary images were constructed from the primary image sets using the FEM-generated DVFs. Slice spacing and image size were kept consistent with the primary image sets with voxel sizes of  $\sim 1$  mm in the X and Y directions, and 3 mm in the Z direction. The FEM-generated DVFs were considered the gold standards to evaluate registrations performed from the primary images to the FEM-simulated images.

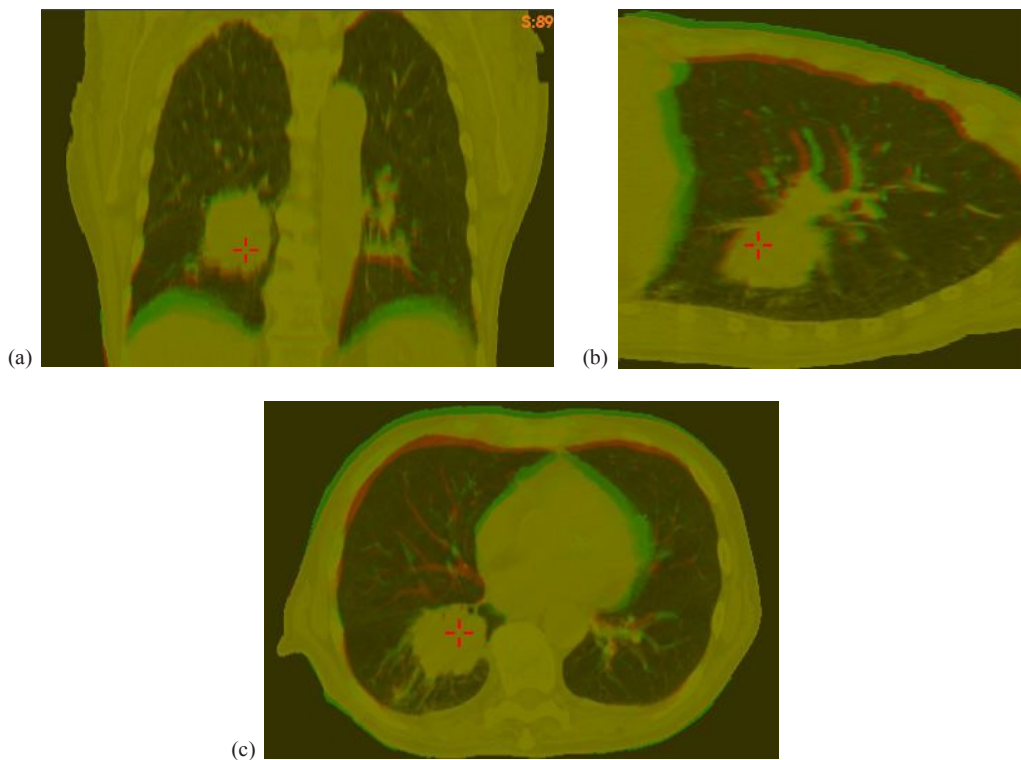


FIG. 1. The original CT image (red) and its deformed image (green) of Lung1<sub>2cm</sub>: (a), (b), and (c) are the coronal, sagittal, and transverse cuts of their overlaid images, respectively.

### **B.2 Computational phantom-based evaluations of DSP with different grid resolutions**

DSP and DMP share the same registration algorithm with their difference mainly in that DSP allows the user to choose the B-spline grid resolution from 1 to 9, with 1 being the coarsest and 9 being the finest. The exact grid spacing of each level is unknown due to the proprietary nature of the software. One can compose a set of DSP registrations to form a multiresolution registration chain. For example, images would be registered first using the coarsest setting, and then that result would be fed to the next stage of finer grid spacing.

In order to evaluate the DSP registration algorithm, chains of registrations with different grid resolutions were performed on the two lung patients (Lung1 and Lung2) with four diaphragm deformations (1.8 cm, 2.0 cm, 2.7, and 3.0 cm). For each dataset, six representative compositions of typical grid resolutions were investigated, including three single grid resolutions of 1, 5, and 9, as well as three resolution compositions of 1→5, 1→9, and 1→5→9. The resulting DVFs were compared against the corresponding FEM-generated gold standard DVFs to estimate the registration errors at each image voxel. The DMP algorithm was also applied and evaluated using the same dataset for comparison.

### **B.3 Computational phantom-based evaluation of DMP and Demons**

To evaluate the DMP algorithm which was suggested as the default by the manufacturer, the performances of DMP and Demons registrations were compared based on the developed FEM models. Both DMP and Demons in VelocityAI had no customizable grid settings and were run with the default parameters. Their registrations were performed by deforming the primary image sets to match the secondary image sets. The registration error,  $e$ , at a point  $x$  was quantified by:

$$e(x) = |DVF_{FEM}(x) - DVF_{DIR}(x)| \quad (1)$$

where the first and second terms are the DVFs generated by the FEM modeling and DIR algorithms, respectively.

Both DMP and Demons registrations were performed on seven deformable models, developed from Lung1 and Lung2 to identify the potential trend of registration errors that are associated with different displacement magnitudes; then the two algorithms were evaluated with models developed from four different patients to detect any patient-specific errors. Registration errors are generally large in the superior and inferior ends of image borders because there is often no corresponding anatomy between two input images. However, the effect is clinically insignificant since the treatment lesions and relevant organs of interest are typically located near the center of the images. Thus, such border regions (2 cm on each end) were excluded from analysis in this study. Boolean masks of the patient body and lung region were created to confine the error analysis to these relevant voxels.

### **B.4 Physical phantom-based evaluation of DMP and Demons**

A motor-controlled lung phantom was developed to simulate respiration-induced deformation. Phantom characterization is beyond the scope of this work; but briefly, the insert was made of heterogeneous sponge with average density equivalent to lung density in CT images. The inferior end of the phantom was deformed 2.5 cm to simulate a diaphragm movement. A 4D CT dataset including the phases of the end inhalation (EI) and end exhalation (EE) was acquired for this phantom. Registrations were performed from EI to EE using both DMP and Demons algorithms. Registration errors were quantified with the ROIs set to the lung region, as well as to the entire image volume. Further, the influence of different window/level settings was investigated. The selected window settings included 250, 500, 1000, and full range with the levels set to the half of the corresponding window values.



The registrations of the phantom images were evaluated with 37 landmarks. These landmarks were sponge features automatically identified in the phantom's EE image. Specifically, the image was scanned voxel by voxel until one was found with intensity greater than a given threshold. Once this voxel was found, a subprocess utilized a spanning tree algorithm<sup>(37)</sup> to find all the directly neighboring voxels with a new threshold that had lower intensity than the first. The first threshold ensured that only objects whose intensity rose above it would be chosen. The second threshold was set lower to include more periphery voxels and form a complete feature. Once the feature was delineated, its center of mass was determined to define the feature's spatial location. All voxels within an already defined object were then ignored in the remaining scans.

A deformable image registration was performed from the EI image to the EE image, and its resultant DVF was then used to construct a warped image from the EI image. The process used to identify the features on the EE image was repeated on this warped image. At this point, two sets of features were determined: one on the EE and another on the warped EI images. The correspondence of these features was established based on their size and center-of-mass coordinates, and was then visually reviewed with their overlaid images. The distance between the corresponding center-of-mass coordinates in the EE image and in the warped EI image was determined, indicating the displacement error of the registration at the center.

## C. Evaluation of image registration algorithms with prostate CT images

### C.1 Development of computational prostate phantoms

The primary CT image sets were acquired from prostate cancer patients using a retrospective protocol approved by the IRB of our institution. One image set with an empty bladder was selected to create a deformable prostate phantom, where bladder expansion was simulated through equally distributed forces applied outward on a set of tetrahedral nodes located in the inside of the bladder. Two different forces were added to these boundary nodes to create two deformed image sets, termed  $P_1^{CT}$  and  $P_2^{CT}$ , resulting in a 6 mm and an 8 mm isotropic expansion of the bladder radius in the lateral direction. The resultant displacements of all the tetrahedral nodes were interpolated to get displacements in the entire image domain. The interpolated displacement fields were used to deform the primary image to generate deformable prostate phantoms with known deformation fields.

### C.2 Prostate phantoms for evaluation of image registrations between CT and CBCT

To evaluate DMP performance between CT and CBCT image registrations, simulated CBCT images were created by incorporating simulated CBCT noise into the FEM-generated benchmark models. Specifically, with the two FEM-simulated deformation fields, the primary prostate CT image was warped to create two deformed images. To create simulated CBCT images,  $P_i^{Noise}$  ( $i = 1, 2$ ), Poisson noise was incorporated into the deformed images in a manner similar to Murphy et al.<sup>(38)</sup> Specifically, the deformed CT image was transformed to its spectrum space using a Fast Fourier Transform (FFT) in MATLAB (The MathWorks, Inc., Natick, MA). Based on Jaffray and Siewerdsen's study of CBCT performance characteristics,<sup>(39)</sup> the noise power spectrum (NPS) was approximated by the function

$$p(k) = e^{-k\lambda}(1 - e^{-k\lambda}) \quad (2)$$

with  $\lambda = 0.15$  mm, where  $k$  is the wave number in the Fourier domain.<sup>(38)</sup> The FFT image was then multiplied by the NPS and converted back to the image domain using an inverse FFT. The resultant image  $P_i^{Noise}$  contains the simulated CBCT noise. If a registration performed from the primary prostate CT image to  $P_i^{CT}$  or  $P_i^{Noise}$  is accurate, its DVF should be equal to the FEM-generated deformation field.

### C.3 Evaluation of registrations between prostate CT and CBCT images

The evaluation of prostate image registrations was performed in the same way as for the lung cases (i.e., a voxel-by-voxel comparison to the FEM simulated DVF for constructed image phantoms). This was done for registrations from the primary CT to  $P_i^{Noise}$  and  $P_i^{Noise}$  ( $i = 1, 2$ , respectively), where  $P_i^{Noise}$  was used to assess the impact of CBCT noise on image registrations. For this reason, average errors and displacement profiles of these  $P_i^{Noise}$  registrations were compared to the  $P_i^{CT}$  registrations. The calculated errors were averaged in the entire patient volume, as well as in the prostate, bladder, and rectum regions.

## III. RESULTS

### A. Evaluation of image registrations using lung CT phantoms

#### A.1 Computational lung phantoms

Figure 1 demonstrates an example for the case of Lung1<sub>2cm</sub> where the center of the diaphragm surface was moved up by 2 cm and the other internal structures were deformed by the FEM model. The overlay of the original image (red) and the FEM deformed (green) image was shown in Figs. 1(a) – 1(c). Similarly, computational phantoms were developed for the four lung cancer patients with a set of different deformation magnitudes.

#### A.2 Evaluation of DSP with different grid settings

DSP registrations with different B-spline grid settings were performed on the computational phantoms Lung1<sub>2cm</sub>, Lung1<sub>3cm</sub>, Lung2<sub>1.8cm</sub>, and Lung2<sub>2.7cm</sub>. All the grid settings defined in the Materials & Methods section B.2 (above) were tested, and the resultant displacement errors were averaged in the patient volume with the results shown in Fig. 2.

As shown in Fig. 2, DMP outperformed DSP with these selected settings for Lung1<sub>2cm</sub>, while a slight improvement of DSP over the DMP algorithm was observed in Lung2<sub>1.8cm</sub> where the average displacement error was improved by 0.2 mm. For the cases of large diaphragm motions, Lung1<sub>3cm</sub> and Lung2<sub>2.7cm</sub>, the DMP performed better for all six settings, and in some cases the differences were larger than 3.0 mm. In general, the performance of the DMP was comparable or superior to that of the DSP with the selected B-spline grid settings. It is noticeable that, for DSP, the  $R_1$  resolution registrations have less error than  $R_9$  registrations, which is particularly true for large deformation cases Lung1<sub>3cm</sub> and Lung2<sub>2.7cm</sub>. It is primarily due to

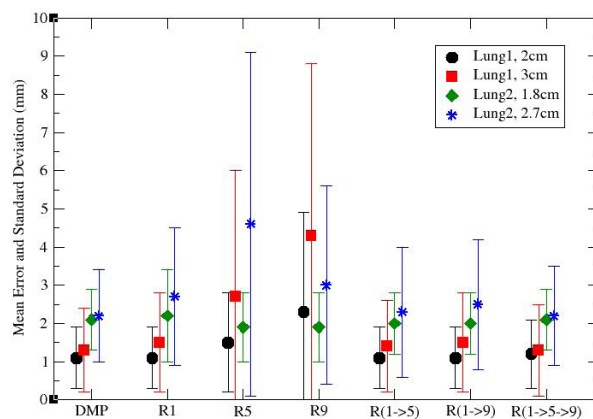


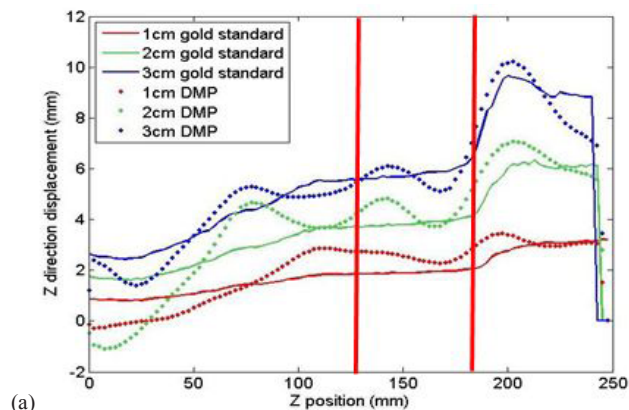
FIG. 2. Average displacement errors ( $\pm$  standard deviation) for different B-spline grid settings.  $R_x$  denotes a DSP registration with the grid resolution  $x$ , and  $R(x \rightarrow y)$  represents the composition of the DSP registrations with the resolution chain  $x \rightarrow y$ .

the nature of the local optimization algorithms, where the optimization is trapped in one of the local minima. Therefore, it is suggested that low resolution be used, especially in the presence of large deformation, followed by gradually finer resolutions for detailed anatomy alignments. This may suggest that grid resolutions should be adapted for different patient images, as well as different deformation scenarios.

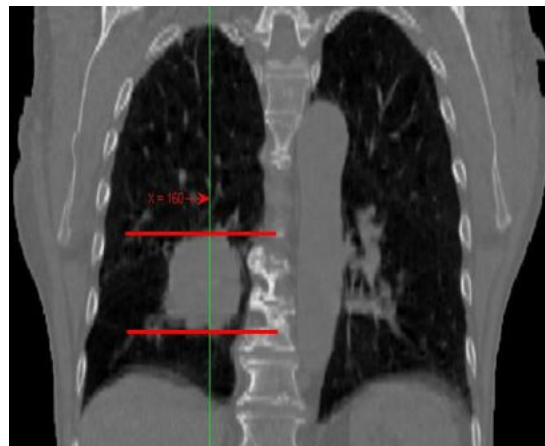
### A.3 Evaluation of DMP and Demons registrations

DMP registrations were performed on the computational phantoms developed from Lung1 and Lung2 at different deformation magnitudes. Figure 3(a) shows the profiles for the superior–inferior (SI) components of the displacement vectors derived from the registrations of Lung1. These profiles correspond to the SI-line in Fig. 3(b) that passes through the tumor volume of Lung1. The position  $z = 0$  mm indicates the most superior slice of the image. It is evident that the DMP displacements are dipping above and below the gold standard sinusoidally due to the B-spline interpolation of the DMP algorithm.

The registration errors in the lung as well as in the patient body for DMP and Demons algorithms were summarized in Table 1. It was observed that the mean displacement error was larger for larger diaphragm motion. In all seven cases, the Demons registrations are slightly



(a)



(b)

FIG. 3. Superior–inferior (SI) displacement profiles derived from DMP registrations for Lung1: (a) the profiles contain displacements from the gold standard DVFs and DMP DVFs at diaphragm deformation of 1, 2, and 3 cm; (b) the SI line illustrated in Lung1's primary CT image is corresponding to the displacement profiles in (a). The tumor region is marked with two red lines on both the image and profile figures.



TABLE 1. Average displacement errors (mm) and standard deviation for DMP and Demons registrations evaluated with computational phantoms of different deformation magnitudes.

<i>Image Set (motion magnitude)</i>	<i>Patient Volume</i>		<i>Lung</i>	
	<i>DMP</i>	<i>Demons</i>	<i>DMP</i>	<i>Demons</i>
Lung1 (10 mm)	1.0±0.7	0.8±0.5	0.9±0.5	0.7±0.4
Lung1 (20 mm)	1.1±0.8	0.9±0.7	0.9±0.6	0.7±0.4
Lung1 (30 mm)	1.3±1.1	1.1±1.0	1.0±0.7	0.8±0.6
Lung1 (40 mm)	1.6±1.5	1.4±1.6	1.2±1.1	1.1±1.2
Lung2 (18 mm)	2.1±0.8	5.2±5.7	2.1±0.8	2.1±0.9
Lung2 (27 mm)	2.2±1.2	5.2±5.4	2.4±1.1	2.3±1.2
Lung2 (42 mm)	3.1±1.7	5.9±5.6	3.3±2.1	3.0±1.4

better than DMP inside the lung, but are much worse outside the lung region for Lung2. The patient-to-patient differences are illustrated further in the next section.

#### **A.4 Evaluation of DMP and Demons registrations with different patients and regions of interest**

Four patients were simulated with their diaphragm deformed about 3 cm in magnitude to create computational phantoms, and both the DMP and Demons algorithms were applied to these phantoms. The mean errors for the DMP registrations are between 1.3 and 2.6 mm, but the performance of the Demons registrations varies largely among these patients (Table 2).

Compared to DMP, Demons has larger errors on the chest wall and breast regions. Inside the lung, Demons has a slightly better performance than DMP in all the cases listed in Tables 1 and 2 except Lung4 (Fig. 4), where a large amount of homogeneous tissues present a major challenge to the Demons registrations (Fig. 4(c)).

Even for one patient, different regions could be registered with different qualities. As shown in Figs. 5(a) and 5(b), errors from DMP and Demons registrations are overlaid on top of the corresponding image for Lung2<sub>2.7cm</sub>, where large errors from the Demons registration can be observed on chest wall. Similarly, Figs. 5(c) and 5(d) show the errors of the two registration algorithms overlaid with the coronal cuts of the primary CT image for Lung1<sub>3cm</sub>. Both DMP and Demons have large errors in the regions near the lateral chest wall and the diaphragm as pointed by the arrows in Figs. 5(c) and 5(d). These images show the effect of large volumes of homogenous tissue on the respective algorithms.

Overall, the average displacement errors of the DMP and Demons registrations in the lung region ranged from 1 mm to 3.3 mm. In all cases except one, the DMP algorithm produced slightly worse registrations than the Demons algorithm in the lung, but it still provided DVFs with comparable accuracy. There were notable differences between the DMP and Demons registrations outside of the lung region. The Lung2 case demonstrated significantly reduced displacement errors for both the breast tissue and chest wall region when the DMP algorithm was used. Here, displacement errors exceeded 4 cm in the breast tissue for the Demons algorithm, but were only 1–2 mm for DMP in the same area.

TABLE 2. Average displacement errors (mm) and standard deviations for DMP and Demons registrations evaluated with different patient images.

<i>Image Set (motion magnitude)</i>	<i>Patient Volume</i>		<i>Lung</i>	
	<i>DMP</i>	<i>Demons</i>	<i>DMP</i>	<i>Demons</i>
Lung1 (30 mm)	1.3±1.1	1.1±1.0	1.0±0.7	0.8±0.6
Lung2 (27 mm)	2.2±1.2	5.2±5.4	2.4±1.1	2.3±1.2
Lung3 (30 mm)	2.6±1.7	2.1±1.3	2.4±1.3	1.9±0.7
Lung4 (30 mm)	2.4±2.3	30.0±18.1	1.7±1.0	9.3±6.3

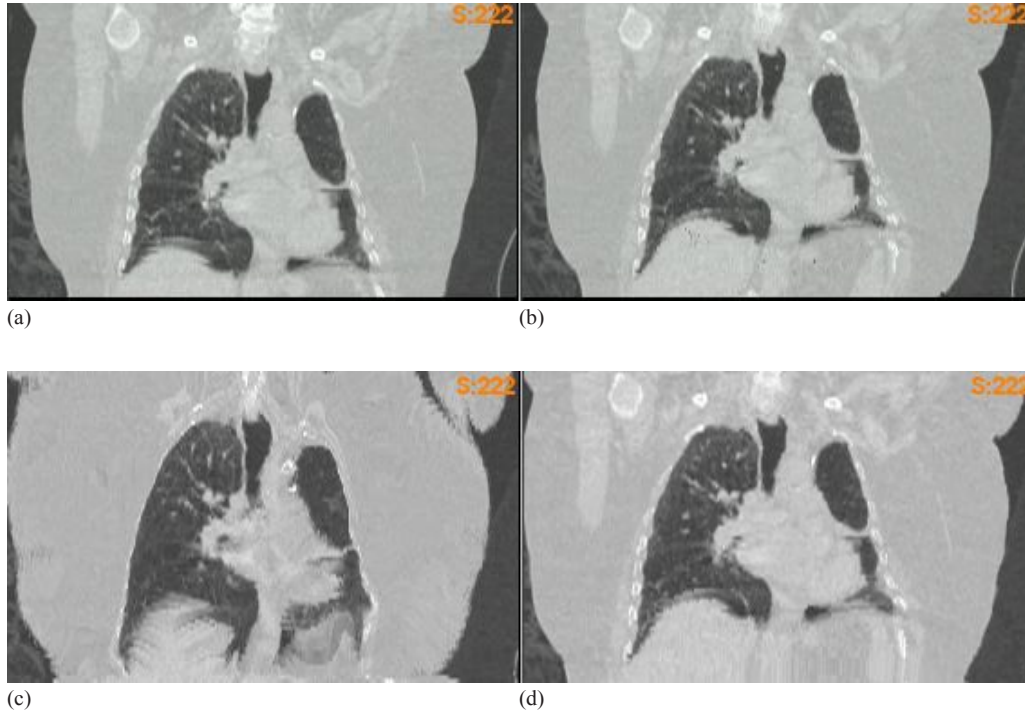


FIG. 4. Computational lung phantom Lung4 with large areas of homogeneous tissue: (a) the original image set; (b) the FEM simulated image set; (c) image warped by Demons DVF, and (d) image warped by DMP DVF.

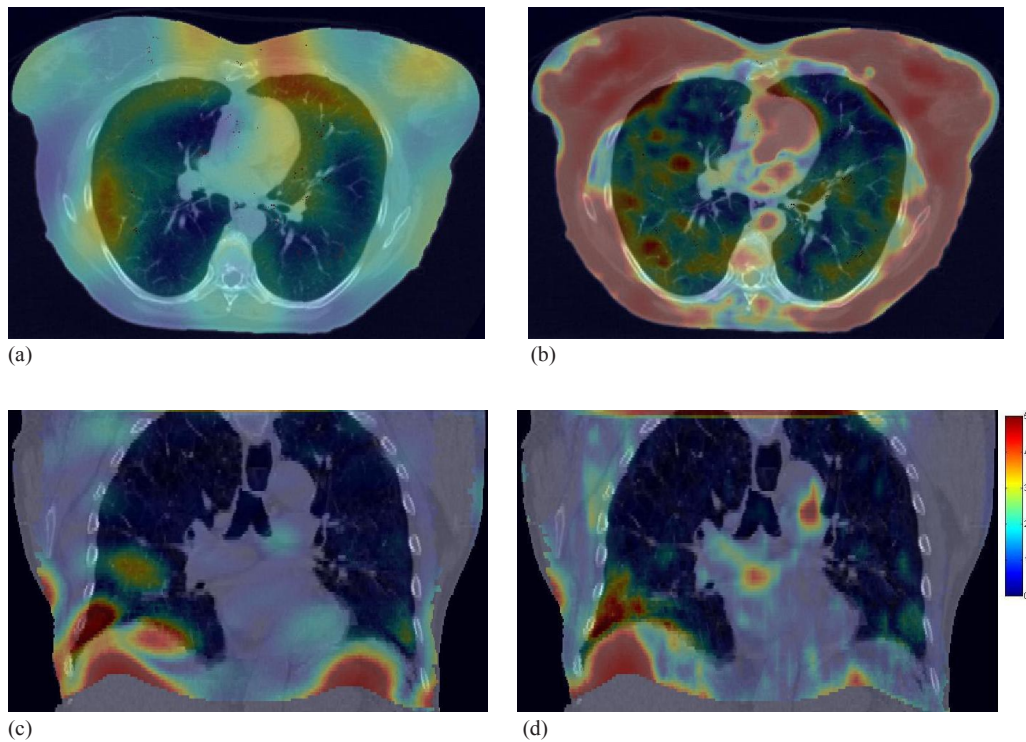


FIG. 5. Color overlay of displacement errors with CT image sets. The figure contains an axial cut of image set Lung2<sub>2.7cm</sub> overlaid with the errors of (a) DMP registration and (b) Demons registration. Also shown is a coronal slice of Lung1<sub>3cm</sub> overlaid with (c) DMP and (d) Demons registration errors. Error values seen in the color bar (d) are given in mm.

### B. Evaluation of DMP and Demons algorithms with a physical phantom

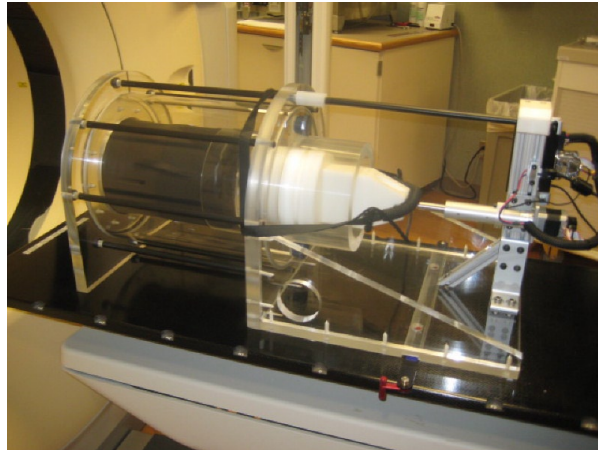
DMP and Demons registrations were performed from EI to EE for the physical phantom. The centers of the 37 previously defined landmark objects were automatically tracked. The spatial difference between the centers in the EE image and those in the warped EI image was counted as the registration error. When the registrations were performed over the entire image volume, their average displacement errors were 1.9 and 1.6 mm for the DMP and Demons algorithms, respectively. However, when restricted to the lung regions, the Demons algorithm failed to produce a visually acceptable registration, resulting in more than 10 mm mean error, while the DMP had its average displacement error reduced to 1.8 mm.

Note that the default parameters in DMP were set for patient images which were largely different from the physical phantom. To address this issue, the image intensity window was limited during the DMP registration, based on the manufacturer's suggestion. Table 3 summarizes average displacement errors for registrations at different window intensity settings. The average of the displacement errors for all the registrations that used customized contrast settings was  $0.97 \pm 0.51$  for the lung region, and  $1.25 \pm 0.58$  for the image domain. The best registrations were in the ROI-constrained registrations with lower window settings,  $R_{250}^{ROI}$  and  $R_{500}^{ROI}$ .

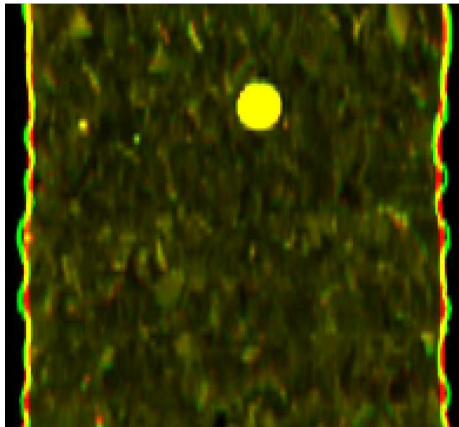
Figure 6 shows the EE image overlaid with the warped EI image using the DMP registrations with the intensity windows of 250 and 500, respectively. The mismatched part of the tumor for the registration  $R_{500}^{ROI}$  can be observed in the overlaid images. The results show that the intensity window could impact the DMP registrations, and is worth investigating further in future studies.

TABLE 3. Average displacement errors and standard deviations of DMP registrations.

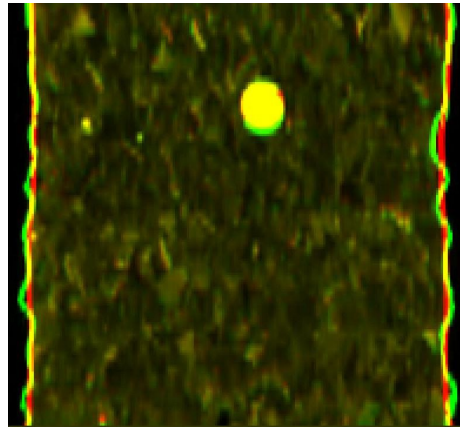
<i>Mean Error (mm)</i>	<i>DMP Registration</i>				
	250	500	1000	1500	Unlimited
Intensity Window					
Image Domain	1.3±0.6	1.3±0.5	1.1±0.5	1.3±0.7	1.9±0.8
Lung Region	0.8±0.4	0.8±0.4	1.2±0.7	1.1±0.5	1.8±0.6



(a)



(b)



(c)

FIG. 6. Physical phantom image sets: (a) the physical phantom; the overlay of EE and warped EI images registered with (b) 250 window setting and (c) 500 window setting. The green image is the EE image and the red one is the warped EI image. The large circular object represents the tumor.

## C. Evaluation of DMP and Demons registrations with prostate CT images

### C.1 Computational prostate phantoms

Figures 7(a) and 7(b) show the original and computationally deformed prostate images. The expansion of the bladder and the deformation of its nearby structures can be observed in the elliptical region marked in the axial and coronal images, respectively. With the method described in the Material and Methods section C.2 (above), simulated CBCT noise was incorporated into the deformed image (Fig. 7(b)) to derive a simulated CBCT image (Fig. 7(c)).

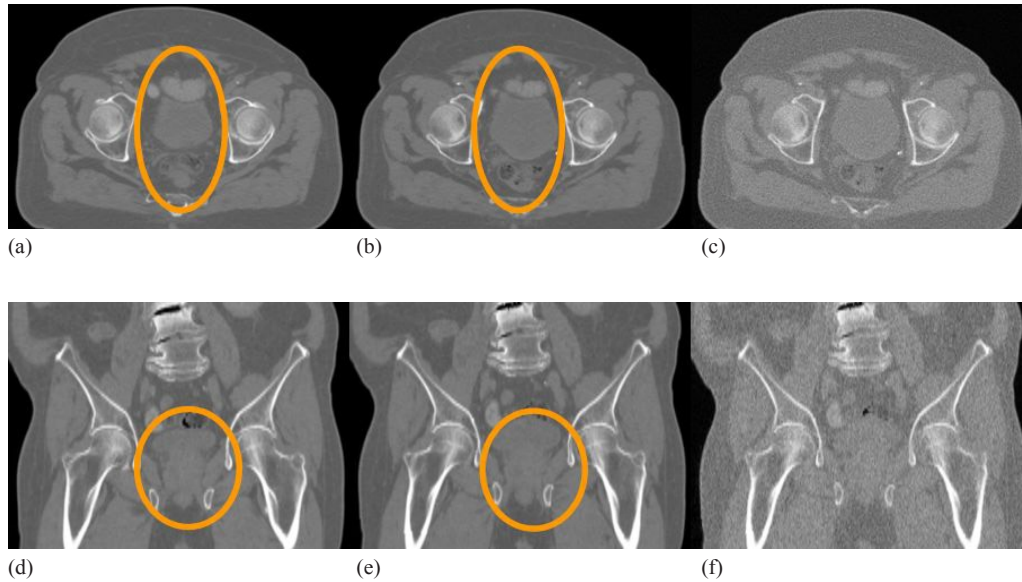


FIG. 7. Computational prostate phantom and simulated CBCT image sets. The axial cuts of (a) the original image set, (b) the deformed image set, and (c) the deformed image with simulated CBCT noise added. The subfigures (d)-(f) are the coronal views of (a)-(c), respectively.



### C.2 Evaluation of DMP and Demons registrations using computational prostate phantoms

Table 4 lists the errors for the DMP and Demons registrations. For DMP, the average registration errors were 1.9 and 2.1 mm across the whole patient volume, and 1.2–2 mm for prostate and rectum. The errors in bladder, 4.6 mm on average, were relatively large primarily due to the large initial deformation with low image gradients in the bladder. For the Demons registrations, without noise added, the displacement errors across the whole patient body were large, 10.3 mm and 7.8 mm for  $P_1^{CT}$  and  $P_2^{CT}$ , respectively (the first and second rows in Table 4). The large errors were mostly observed in the outer tissue regions near the skin surface, while regions of clinical interest, including the prostate and rectum, yielded registration errors between 2.2–3.2 mm.

TABLE 4. Average displacement errors (mm) and standard deviation for DMP and Demons registrations for pelvic images.

Image Set	Whole Patient		Prostate		Rectum		Bladder	
	DMP	Demons	DMP	Demons	DMP	Demons	DMP	Demons
$P_1^{CT}$	2.1±0.7	10.3±11.5	1.9±0.5	3.2±1.2	1.9±0.5	2.2±0.9	2.8±1.5	3.7±1.8
$P_2^{CT}$	1.9±1.0	7.8±8.3	1.2±0.6	2.4±1.3	2.0±0.8	2.4±1.0	4.6±3.1	5.0±3.4
$P_1^{Noise}$	2.1±0.8	18.5±18.1	1.7±0.5	12.3±10.5	1.8±0.6	12.8±11.2	3.1±1.7	16.3±13.9
$P_2^{Noise}$	1.8±0.9	19.4±20.7	1.0±0.6	13.2±11.7	2.4±1.0	13.3±12.0	5.0±3.4	18.3±14.5

For the simulated CBCT images, the Demons algorithm was not able to produce visually satisfactory results and its registration errors were greatly increased. The impact of the simulated noise on DMP was small, with slight variations observed from region to region. Table 4 lists the average displacement errors for all the DMP and Demons registrations.

Figure 8(a) shows the sagittal view of  $P_1^{CT}$ . Figures 8(b) and 8(c) are its corresponding images generated by the DMP and Demons registrations. Their displacement error maps overlaid on  $P_1^{CT}$  are shown in Figs. 8(e) and 8(f). The sagittal views illustrate the errors of these registration algorithms in the homogenous fatty tissue, as well as higher contrast medial tissues.

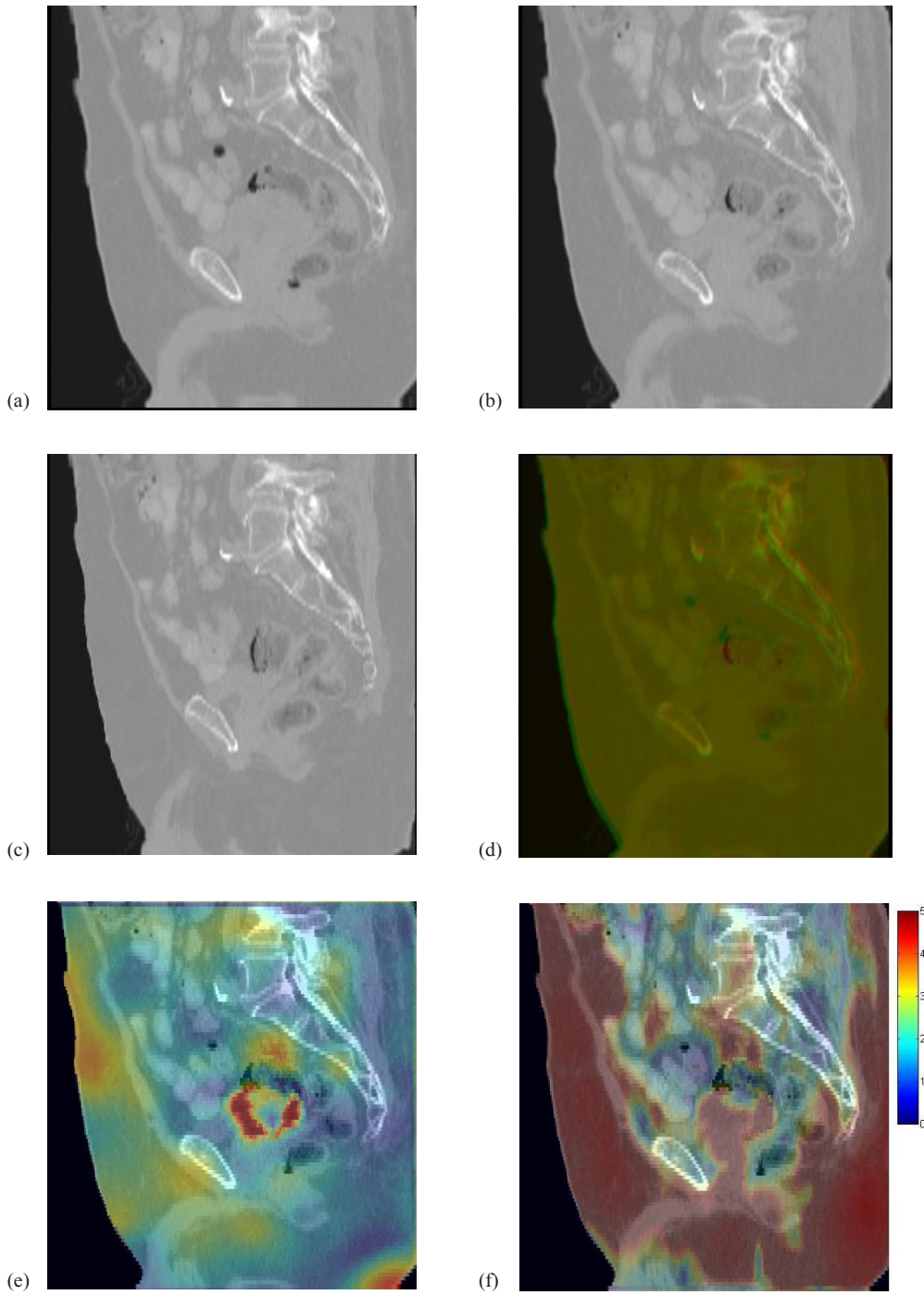


FIG. 8. The FEM constructed image  $P_1^{CT}$  (a); the prostate CT images (b) and (c) warped by DMP-DVF and Demons-DVF; (d) overlay of the warped prostate CT images in (a) and (b); (e) and (f) are the displacement errors (mm) of the DMP registration and the Demons registration overlaid on a sagittal view of the FEM constructed image  $P_1^{CT}$ .

#### IV. DISCUSSION

When adaptive radiation therapy is implemented in clinic, it becomes necessary to quantify the accuracy of its underlying deformable image registrations. In this study, we developed computational phantoms from lung cancer patient CT images to evaluate the performance of the DSP algorithm configured with a variety of B-spline grid setting options. The results revealed that the best performance of the DSP registrations was comparable to DMP. The DMP algorithm had average displacement errors ranging from 1.0 ~ 3.3 mm for the diaphragm compressions of 1 ~ 4 cm. This is comparable to the results of a multi-institution deformable registration accuracy study (MIDRAS).<sup>(16)</sup> In the MIDRAS, the institutions that used the B-spline-based registrations reported 1.6 to 3.0 mm errors when using a set of physician-identified landmarks as gold standards. In contrast to the landmark method, the computational phantom-based evaluation method uses simulated tissue deformation maps as benchmark models, which allow DIRs to be evaluated in a voxel-by-voxel manner, and this method can overcome the limitations of the landmark-based evaluation such as observation uncertainty, lack of image features, or slice thickness-induced round-off errors.

The accuracy of the DMP algorithm was found to be dependent on the magnitude of tissue deformation, as well as individual patient images. As shown in Table 1, the registration errors increased for larger diaphragm deformations. This observation is similar to that reported by Liu et al.<sup>(19)</sup> in which their in-house developed DIR algorithm was evaluated with a computational phantom (NCAT) and a physical liver phantom. However, the capability of their DIR algorithm could be underestimated because the NCAT phantom has uniform intensities assigned to each organ and, therefore, the DIRs performed on this phantom would not be as accurate as those performed on real patients. In contrast, the computational phantoms in this study, derived from actual patient CT images, do not have this limitation. It also further allows registration algorithms to be evaluated under patient-specific clinical scenarios.

With the patient-specific phantoms, the evaluation results showed that DMP registration outperformed Demons outside the lung, but inside the lung the DMP registrations were slightly worse. Both the DMP and Demons registrations demonstrated large errors in the lower lobe of the right lung and at the boundaries between the lung, diaphragm, and chest wall, as marked in Figs. 5(c) and 5(d). At these boundaries, the simulated deformation fields were large, and such large deformation combined with the homogeneity of tissue could be a major source of registration errors.<sup>(40)</sup>

For the prostate cases, DMP had smaller registration errors than Demons in the prostate. For the Demons algorithm, significant errors were observed in regions of low contrast tissue. With the simulated CBCT noise added, the accuracy of the Demons algorithm degraded significantly in all of the structures evaluated in this study, while the DMP algorithm remained robust. The average errors across the prostate, the rectum, and the patient volume indicate that the simulated CBCT noise did not have a major impact on the DMP registration. This confirms the results from Murphy et al.<sup>(38)</sup> where the simulated CBCT noise had no statistically significant impact on the B-spline-based image registrations.

It should be noted that, while different deformation scenarios can be simulated, limitations still exist for the computational phantom-based method. For example, on the 4D lung images it is common that motion artifacts differ among phase images. Furthermore, the differences of the organ filling, such as the rectum feces and gas, between the treatment planning CT and daily CBCT images creates one of the major challenges for deformable registrations. The impact of such artifacts and differences has not yet been investigated. As a supplementary verification, the physical phantom was employed in this study to evaluate these DIR algorithms. However, the deformation of real patients is much more complex than the physical phantom deformation. Patient-specific verifications are still needed for clinical use of these DIR algorithms.

## V. CONCLUSIONS

In this study, patient-specific FEM models have been developed and used as the gold standard to evaluate DIR algorithms implemented in the VelocityAI software package. It has been found that the accuracy of these algorithms is patient-dependent, and related to various factors including the tissue deformation magnitudes and image intensity gradients across the regions of interest. This may suggest that DIR algorithms need to be verified for individual deformation instances when implementing adaptive radiation therapy.

## ACKNOWLEDGMENTS

This study is financially supported by NIH/NCI Grant No. R01CA140341.

## REFERENCES

1. Frazier A, Yan D, Du M, et al. Effects of treatment setup variation on beam's eye view dosimetry for radiation therapy using the multileaf collimator vs. the cerrobend block. *Int J Radiat Oncol Biol Phys.* 1995;33(5):1247–56.
2. Frazier RC, Vicini FA, Sharpe MB, et al. Impact of breathing motion on whole breast radiotherapy: a dosimetric analysis using active breathing control. *Int J Radiat Oncol Biol Phys.* 2004;58(4):1041–47.
3. Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. *Phys Med Biol.* 1997;42(1):123–32.
4. Kessler ML. Image registration and data fusion in radiation therapy. *Br J Radiol.* 2006;79 Spec No 1:S99–S108.
5. Jenkinson M and Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal.* 2001;5(2):143–56.
6. Rohr K, Stiehl HS, Sprengel R, Buzug TM, Weese J, Kuhn MH. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Trans Med Imaging.* 2001;20(6):526–34.
7. Rueckert D, Aljabar P, Heckemann RA, Hajnal JV, Hammers A. Diffeomorphic registration using B-splines. *Med Image Comput Assist Interv.* 2006;9(Pt 2):702–09.
8. Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal.* 1998;2(3):243–60.
9. Silva JS, Cancela J, Teixeira L. Fast volumetric registration method for tumor follow-up in pulmonary CT exams. *J Appl Clin Med Phys.* 2011;12(2):362–65.
10. Glocker B, Paragios N, Komodakis N, Tziritas G, Navab N. Inter and intra-modal deformable registration: continuous deformations meet efficient optimal linear programming. *Inf Process Med Imaging.* 2007;20:408–20.
11. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging.* 1999;18(8):712–21.
12. Lu W, Chen ML, Olivera GH, Ruchala KJ, Mackie TR. Fast free-form deformable registration via calculus of variations. *Phys Med Biol.* 2004;49(14):3067–87.
13. Fallone BG, Rivest DR, Riauka TA, Murtha AD. Assessment of a commercially available automatic deformable registration system. *J Appl Clin Med Phys.* 2010;11(3):3175.
14. Bender ET and Tome WA. The utilization of consistency metrics for error analysis in deformable image registration. *Phys Med Biol.* 2009;54(18):5561–77.
15. Hardcastle N, Tome WA, Cannon DM, et al. A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy. *Radiat Oncol.* 2012;7:90.
16. Brock KK. Results of a multi-institution deformable registration accuracy study (MIDRAS). *Int J Radiat Oncol Biol Phys.* 2010;76(2):583–96.
17. Castillo R, Castillo E, Guerra R, et al. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys Med Biol.* 2009;54(7):1849–70.
18. Kashani R, Hub M, Balter JM, et al. Objective assessment of deformable image registration in radiotherapy: a multi-institution study. *Med Phys.* 2008;35(12):5944–53.
19. Liu F, Hu Y, Zhang Q, Kincaid R, Goodman KA, Mageras GS. Evaluation of deformable image registration and a motion model in CT images with limited features. *Phys Med Biol.* 2012;57(9):2539–54.
20. Vaman C, Staub D, Williamson J, Murphy MJ. A method to map errors in the deformable registration of 4DCT images. *Med Phys.* 2010;37(11):5765–76.
21. Schreiber E, Pantalone P, Waller A, Fox T. A measure to evaluate deformable registration fields in clinical settings. *J Appl Clin Med Phys.* 2012;13(5):126–39.
22. Zhong H, Peters T, Siebers JV. FEM-based evaluation of deformable image registration for radiation therapy. *Phys Med Biol.* 2007;52(16):4721–38.
23. Bender ET and Tome WA. The utilization of consistency metrics for error analysis in deformable image registration. *Phys Med Biol.* 2009;54(18):5561–77.
24. Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage.* 2009;46(3):786–802.

25. Heath E and Seuntjens J. A direct voxel tracking method for four-dimensional Monte Carlo dose calculations in deforming anatomy. *Med Phys.* 2006;33(2):434–45.
26. Zhong H and Siebers JV. Monte Carlo dose mapping on deforming anatomy. *Phys Med Biol.* 2009;54(19):5815–30.
27. Salguero FJ, Saleh-Sayah NK, Yan C, Siebers JV. Estimation of three-dimensional intrinsic dosimetric uncertainties resulting from using deformable image registration for dose mapping. *Med Phys.* 2011;38(1):343–53.
28. Lee HP, Lin MC, Foskey M. Physically-based validation of deformable medical image registration. *Med Image Comput Comput Assist Interv.* 2008;11(Pt 2):830–38.
29. Zhong H, Kim J, Chetty IJ. Analysis of deformable image registration accuracy using computational modeling. *Med Phys.* 2010;37(3):970–79.
30. Wang H, Dong L, O’Daniel J, et al. Validation of an accelerated ‘demons’ algorithm for deformable image registration in radiation therapy. *Phys Med Biol.* 2005;50(12):2887–905.
31. Serban M, Heath E, Stroian G, Collins DL, Seuntjens J. A deformable phantom for 4D radiotherapy verification: design and image registration evaluation. *Med Phys.* 2008;35(3):1094–102.
32. Chang J, Suh TS, Lee DS. Development of a deformable lung phantom for the evaluation of deformable registration. *J Appl Clin Med Phys.* 2010;11(1):281–86.
33. Kerdok AE, Cotin SM, Ottensmeyer MP, Galea AM, Howe RD, Dawson SL. Truth cube: establishing physical standards for soft tissue simulation. *Med Image Anal.* 2003;7(3):283–91.
34. Nie K, Chuang C, Kirby N, Braunstein S, Pouliot J. Site-specific deformable imaging registration algorithm selection using patient-based simulated deformations. *Med Phys.* 2013;40(4):041911.
35. 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;14(1):192–213.
36. Zhong H, Jin JY, Ajlouni M, Movsas B, Chetty IJ. Measurement of regional compliance using 4DCT images for assessment of radiation treatment. *Med Phys.* 2011;38(3):1567–78.
37. Nesetril J, Milkova E, Nesetrilova H. Otakar Boruvka on minimum spanning tree problem. Translation of both the 1926 papers, comments, history. *Discrete Mathematics.* 2001;233(1-3):3–36.
38. Murphy MJ, Wei Z, Fatyga M, et al. How does CT image noise affect 3D deformable image registration for image-guided radiotherapy planning? *Med Phys.* 2008;35(3):1145–53.
39. Jaffray DA and Siewerdsen JH. Cone-beam computed tomography with a flat-panel imager: initial performance characterization. *Med Phys.* 2000;27(6):1311–23.
40. Zhong H, Kim J, Li H, Nurushev T, Movsas B, Chetty IJ. A finite element method to correct deformable image registration errors in low-contrast regions. *Phys Med Biol.* 2012;57(11):3499–515.