Analysis of deformable image registration accuracy using computational modeling

Hualiang Zhong,^{a)} Jinkoo Kim, and Indrin J. Chetty

Department of Radiation Oncology, Henry Ford Health System, Detroit, Michigan 48202

(Received 8 June 2009; revised 11 January 2010; accepted for publication 11 January 2010; published 4 February 2010)

Computer aided modeling of anatomic deformation, allowing various techniques and protocols in radiation therapy to be systematically verified and studied, has become increasingly attractive. In this study the potential issues in deformable image registration (DIR) were analyzed based on two numerical phantoms: One, a synthesized, low intensity gradient prostate image, and the other a lung patient's CT image data set. Each phantom was modeled with region-specific material parameters with its deformation solved using a finite element method. The resultant displacements were used to construct a benchmark to quantify the displacement errors of the Demons and B-Spline-based registrations. The results show that the accuracy of these registration algorithms depends on the chosen parameters, the selection of which is closely associated with the intensity gradients of the underlying images. For the Demons algorithm, both single resolution (SR) and multiresolution (MR) registrations required approximately 300 iterations to reach an accuracy of 1.4 mm mean error in the lung patient's CT image (and 0.7 mm mean error averaged in the lung only). For the low gradient prostate phantom, these algorithms (both SR and MR) required at least 1600 iterations to reduce their mean errors to 2 mm. For the B-Spline algorithms, best performance (mean errors of 1.9 mm for SR and 1.6 mm for MR, respectively) on the low gradient prostate was achieved using five grid nodes in each direction. Adding more grid nodes resulted in larger errors. For the lung patient's CT data set, the B-Spline registrations required ten grid nodes in each direction for highest accuracy (1.4 mm for SR and 1.5 mm for MR). The numbers of iterations or grid nodes required for optimal registrations depended on the intensity gradients of the underlying images. In summary, the performance of the Demons and B-Spline registrations have been quantitatively evaluated using numerical phantoms. The results show that parameter selection for optimal accuracy is closely related to the intensity gradients of the underlying images. Also, the result that the DIR algorithms produce much lower errors in heterogeneous lung regions relative to homogeneous (low intensity gradient) regions, suggests that feature-based evaluation of deformable image registration accuracy must be viewed cautiously. © 2010 American Association of Physicists in Medicine.

[DOI: [10.1118/1.3302141](https://doi.org/10.1118/1.3302141)]

Key words: finite element modeling, deformable registration, adaptive radiotherapy, quality assurance

I. INTRODUCTION

Deformable image registration (DIR) is an integral component of the image-guided radiation therapy (IGRT) and image-guided adaptive radiation therapy processes.^{1,2} Specific applications include automatic contouring,^{3,4} image synthesis for dose calculation,⁵ and dose accumulation for adaptive planning.^{6–8} An important aspect of DIR is the validation process, which is necessary in the development of 4D or adaptive radiotherapy techniques, which deliver non-uniform dose distributions to different parts of the target volume and organs at risk. A clinical workflow recently described by Yan² for adaptive radiotherapy highlighted the need for quality assurance of DIR. Due to lack of gold standards in clinical settings, validation of DIR is often challenging. Consequently, various validation methods have been proposed to address different scenarios.

For example, inspecting difference images is convenient for daily clinical verification and may provide a qualitative

evaluation of the deformation accuracy.⁹ Contour-based evaluation may provide an assessment of the deformation map which is especially useful for contour propagation,^{10,11} and landmark points within regions of interest may convey valuable information for quality assurance.^{9,11,12} In addition, the concept of self or inverse consistency may allow the accuracy of deformation registration to be evaluated.^{13–16} Computational phantoms offer another option to directly derive the displacement errors at each voxel. Built from tomographic images of real subjects, the phantom deformation can be simulated using different mathematical formulae^{17–21} so that displacements can be compared in submillimeter anatomical detail. These phantoms were used in the investigation of the interplay effect and the effect of dynamic lung density in IMRT treatments.^{22,23} However, mathematic deformations may not be sufficiently realistic in simulation due to the variations in tissue elasticity (soft tissue vs rigid bone), compressibility (rectum vs prostate), and organ discontinuity (diaphragm vs chest wall). To address these deficiencies, ad-

vanced physical phantoms have been developed. For example, Serban *et al.*²⁴ glued nylon wires, lucite beads, and dermasol ellipsoids (emulating vascular, bronchial bifurcations, and tumors) at various locations throughout dampened sponges, with the beads and bifurcations used for the deformable registration and its validation. Kashani *et al.*²⁵ embedded a sufficient number of markers into their phantom for the validation of DIR and dosimetry studies. As part of a multi-institutional study, Kashani *et al.*²⁶ performed a quantitative evaluation of eight different registration models based on the trajectory of the embedded markers in the phantom. Recently, Castillo *et al.*²⁷ used large samples (>1100) of corresponding pulmonary landmark features manually selected from 4D CT data sets for the evaluation of two DIR algorithms and concluded that landmark pairs can be used to assess DIR spatial accuracy with a narrow uncertainty range. These studies presented the comprehensive evaluation of various registration models.

It should be mentioned that while image features such as anatomic structures or bifurcations can be used for quantifying registration errors in the lung, DIR performance in other regions, particularly in regions with low intensity gradients, is still difficult to assess. Furthermore, for image intensity based registration algorithms (such as Demons or B-Spline), DIR errors are prone to appear in regions with low image gradients (as demonstrated in this study). While this type of error can be circumvented using finite element modeling (FEM)-based registrations, which rely on the conservation of elastic energy instead of an image similarity metric,^{12,28,29} the FEM registrations require organ-specific mesh and accurate boundary constraints. It has also been demonstrated³⁰ that computational models without specification of tumor's properties could result in large displacement errors, depending on the size of the tumor.

In this study we developed a FEM technique to simulate the deformation of a synthesized prostate phantom and a lung patient's CT image. The resultant displacements were employed to create the deformed images which were combined with the original images to form a testing framework for validation of different registration algorithms. Specifically, the accuracies of Demons and B-Spline algorithms configured with different parameters were analyzed within this framework, and the detected registration errors were then verified using the concept of unbalanced energy (UE) proposed by Zhong *et al.*³¹

II. MATERIALS AND METHODS

In this study, the finite element modeling technique developed previously³² was integrated with CT images to create two deformable models: A synthesized prostate model and a CT image-based lung model. These models were used to generate displacement vector fields (DVF), which were subsequently employed to synthesize deformed images for validation of deformable registration algorithms, where the intensity of these images was represented as CT number.

II.A. Development of FEM framework

In elasticity theory, the dynamic equilibrium of the entire body V is modeled by the following equation:

$$\int \int_S T dS + \int \int \int_V f^b dV = \int \int \int_V \rho \frac{\partial^2 u}{\partial t^2} dV, \quad (1)$$

where ρ is mass density, u represents the displacements of a point under the influence of a given external traction T on the surface S , resulting in the acceleration $\partial^2 u / \partial t^2$ and the body force f . Based on the constitutive stress-strain relationship that involves two elastic parameters, namely, Young's modulus (E) and Poisson's ratio (ν), the standard formulation procedure may generate a set of linear algebraic equations

$$K \bar{U}^T = \bar{F}^T, \quad (2)$$

where $\bar{U}$ and $\bar{F}^T$ are the displacement and force vectors at the vertices of the discrete mesh, and K is the global stiffness matrix.³² In this study, a volumetric tetrahedral mesh was developed from an image domain. Different elasticity parameters were assigned based on the tissue types identified from their image intensity values. To achieve high computational accuracy, a tetrahedral mesh that consisted of 132 000 nodes and 780 000 tetrahedra was generated by ABAQUS® (ABAQUS, Inc., Providence, RI). This resulted in a global matrix K consisting of 150×10^9 float-point entries. Due to the fact that K is sparse, all the zero entries in the matrix were suppressed in its implementation. The compressed linear equations were then solved using a preconditioned conjugated gradient method described in Barret *et al.*³³

From the displacement vector U_n of the node v_n solved from Eq. (2), the displacement vector of the reference image voxel s was then calculated by the weighed interpolation

$$X_s = \sum_{k=1,2,3,4} w_{s,n_k} U_{n_k}, \quad (3)$$

where U_{n_k} , $k=1 \dots, 4$, is the displacement vector of the vertex v_{n_k} of the tetrahedron that covers the center c_s of the voxel s , and w_{s,n_k} is the volume coordinate defined by Zienkiewicz *et al.*³⁴

$$w_{s,n_k} = \frac{V(S_{n_k}, c_s)}{V(S_{n_1}, c_s) + V(S_{n_2}, c_s) + V(S_{n_3}, c_s) + V(S_{n_4}, c_s)},$$

$$k = 1, \dots, 4, \quad (4)$$

where S_{n_k} represents the triangle face opposite to v_{n_k} and $V(S_{n_k}, c_s)$ is the volume of the tetrahedron (S_{n_k}, c_s) . The set of X_s from all S in the image domain forms a DVF. Later Eq. (3) and (4) will be used to construct a benchmark displacement field which is termed M -DVF (see Sec. II C).

II.B. Development of numerical phantoms

With the developed FEM framework, two numerical phantoms were developed. One is a fabricated prostate phantom and the other is simulated from a patient's CT image. The image of the prostate phantom was fabricated using

FIG. 1. The M -DVF $d_M(\tau)$ constructed from the FEM interpolated displacements $d(p_j)$.

$160 \times 160 \times 160$ voxels, each with size $1 \times 1 \times 1$ mm³. The image contains several basic geometrical shapes representing the bladder, prostate, rectum, and femoral head structures. The femoral heads are represented as two rigid cubes, and other organs are represented as soft ellipsoids in various sizes. A 2 cm depth bar was created and attached to the bladder to illustrate the effect of deformation. Each structure in the image was assigned different CT numbers to create gradients in image intensity at structure boundaries. For consistency with the image domain, the previously partitioned tetrahedral mesh was scaled to cover the $16 \times 16 \times 16$ cm³ cube. The Young's modulus was set to 10 MPa for femoral heads and 10.0 kPa for soft tissue. The Poisson ratio was set to 0.32 for bladder, 0.49 for femoral heads, and 0.45 for other soft tissues. The rigid bones were fixed and external forces were exerted at the center of the bar (attached to the bladder) to induce a desired deformation.

The other deformable phantom developed in this study was derived from the CT data set of a lung cancer patient. It includes $410 \times 260 \times 100$ voxels, each with size $1.0 \times 1.0 \times 2.5$ mm³. The previous mesh was scaled to fit the CT image domain. The new mesh was configured such that the left and right chest walls and spine are fixed. External forces were exerted on the diaphragm. 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. To simulate slippage conditions at the lung boundary, boundary elements (identified using image gradients) were assigned a low Young's modulus (0.1 kPa) to reduce restriction of lung deformation at the chest wall.

II.C. Procedure for validation of deformable image registration

With a computational model developed from a reference image R , a finite element simulation was performed first to compute the displacement at each vertex v_i in the tetrahedral mesh (see Fig. 1). Each tetrahedron is then partitioned with N sampling points p_j . The displacement $d(p_j)$ at each p_j was calculated using Eqs. (3) and (4). From the coordinates of p_j and $d(p_j)$, the index of the target voxel τ was calculated by $\tau = h(p_j + d(p_j))$, where the function h converts the coordinate back to its image index. A counting index k_τ is updated by $k_\tau = k_\tau + 1$. Consequently, $d_M(\tau)$ is updated by $d_M(\tau) = k_\tau - 1/k_\tau d_M(\tau) - 1/k_\tau d(p_j)$. This process was applied to all the

FIG. 2. Diagram for the construction of the simulation image S and the benchmark correlation between S and R .

points p_j . When N is sufficiently large, each voxel τ in the target domain S will be associated with at least one sampling point p_j , except at some boundary voxels, where $d_M(\tau)$ was assigned by $d_M(\tau) = -X_\tau$ and X_τ is calculated directly from Eqs. (3) and (4). This generated an M -DVF $\tilde{d}_M$ from S to R . The procedure for inverse displacements was implemented in the FEM framework. A similar technique was also implemented in the Monte Carlo code for mass reconstruction.³⁵

With the model generated $\tilde{d}_M$, the voxel intensity in the simulation image S can be assigned on a voxel-by-voxel basis, from a reference image R without image interpolation (Fig. 2), i.e., given a voxel τ in S , the image value at τ was assigned by $I_S(\tau) := I_R(\tau_1)$, $\tau_1 = h(g(\tau) + \tilde{d}_M(\tau))$, where the function g converts an image index to its coordinate, and h converts the coordinate back to its image index. The simulation and reference images combined with the M -DVF $\tilde{d}_M$ form a framework for validation of an arbitrary DIR registration.

The performance of an image registration algorithm can be evaluated in two ways. For the algorithms implemented in ITK, if a registration is performed from R to S , its transformation $T_{S \rightarrow R}$ is a spatial mapping from S to R , i.e., $T_{S \rightarrow R}(\tau) = \tau + h(\tilde{d}_{S \rightarrow R}(\tau))$. In this case, the resultant DVF should satisfy the equation $\tilde{d}_{S \rightarrow R}(\tau) = \tilde{d}_M(\tau)$. Let $\|\cdot\|$ denote the Euclidean distance in the simulation image. The displacement error $\varepsilon(\tau)$ can then be defined by $\varepsilon(\tau) = \|\tilde{d}_{S \rightarrow R}(\tau) - \tilde{d}_M(\tau)\|$. This approach is termed a “forward comparison” (Fig. 2). If an image registration is performed from S to R , its transformation mapping $T_{R \rightarrow S}$ is from R to S . The resultant displacements should satisfy that $\tilde{d}_{R \rightarrow S}(\tau_1) = \tilde{d}_{R \rightarrow S}(h(g(\tau) + \tilde{d}_M(\tau))) = -\tilde{d}_M(\tau)$ for any given τ in S . This is called a “reverse comparison” (RC). It should be mentioned that round-off errors of g and h may compromise the accuracy of the RC validation. We will compare the two methods and provide an illustration of their general consistency in evaluation of the influence of different registration parameters.

In the routine clinical environment, the FEM-based benchmark is not usually available. Therefore we have previously developed a more practical approach to quantify DIR displacement errors, based on a correlation of the unbalanced

FIG. 3. The transversal slice of (a) the fabricated prostate image, (b) the deformed image, and (c) their overlay.

energies with the FEM-based model.³¹ This technique is based on the theory that the elasticity energy stored in a given displacement vector field must be balanced by external work. A large value of unbalanced energy in a location indicates the existence of displacement errors in that neighborhood. The unbalanced energy concept will be verified using the FEM-based validation procedure.

II.D. Deformable image registration algorithms

Two image registration algorithms were evaluated in this study, Demons and B-Spline. The demons algorithm³⁶ previously implemented in the ITK software package³⁷ introduces a demons force that is the displacement during the time interval between the two image frames and is defined as

$$v = \begin{cases} \frac{(m-r)(\vec{\nabla}m + \vec{\nabla}r)}{(m-r)^2 + (\vec{\nabla}m + \vec{\nabla}r)^2}, & |m-r| > t_0 \\ 0, & \text{Otherwise} \end{cases}, \quad (5)$$

where $\vec{\nabla}m$ and $\vec{\nabla}r$ represent the moving and reference image gradients, respectively, and the $(m-r)$ term represents the differential intensity between the reference and moving images. The threshold $t_0=0.001$ was set for the normalized intensity difference between the reference and the moving images and the moving image was calculated using B-Spline interpolations. The Demons algorithm was implemented with single resolution (SR) and multiresolution (MR) modes. The gradient tolerance was configured to 0.0001 for the SR-Demons and to [0.005, 0.005, 0.005, 0.0001] for the MR-Demons. A low tolerance was chosen to ensure that the maximum number of iteration loops can be reached in the following tests. The standard deviation of the Gaussian smoothing function was set to 1.0.

In the B-Spline registration,³⁸ the transformation of a point is computed from the cubic B-Spline basis functions using the positions of the surrounding control points which were updated so that the difference between the moving image and target image can be minimized. In this study, the SR-B-Spline and MR-B-Spline algorithms were developed based on the ITK software package. The mean square differ-

ence of image intensities was set as the similarity metric. The maximum number of their iteration loops was set to 500 for the two B-Spline algorithms.

III. RESULTS

III.A. Deformed numerical phantoms

Based on the material property and boundary conditions described in Sec II B, the prostate phantom [Fig. 3(a)] was then deformed using the finite element technique described in Sec II A. The displacements of these vertices were interpolated to generate an M -DVF for the entire prostate image domain, as described in Sec II C. The deformed image reconstructed from the fabricated source image using the M -DVF is shown in Fig. 3(b). The overlay of the two images [Fig. 3(c)] shows that the ellipsoids are deformed due to the simulated compression, but the rigid femoral heads at the lateral sides are undeformed, as expected. The largest ellipsoid representing bladder has a change in volume following the deformation.

Similarly, for the lung patient's CT image [Fig. 4(a)], the displacement variables of the underlying mesh were solved from Eq. (2) based on the parameters and boundary conditions assigned in Sec II B. These displacements were interpolated to generate the M -DVF for the CT image domain. Using the M -DVF, the deformed image was reconstructed from the original reference CT image with the result shown in Fig. 4(b). The overlay of the original and deformed CT images is shown in Fig. 4(c). From the overlay, it is observed that the bony structures (spine and ribs) remain rigid, but the diaphragm (which has a 2.6 cm peak-to-peak movement) caused deformation of the lung and the tumor. External forces exerted on the diaphragm caused motion primarily in the superior/inferior direction, and the center of the tumor located at the voxel [158, 175, 62] was shifted 0.6 cm in the superior direction.

Since the deformed image [Fig. 4(b)] was constructed from the reference image [Fig. 4(a)] using the M -DVF, the

FIG. 4. The coronal slice of (a) the original lung CT image, (b) CT image deformed with FEM displacements, and (c) the overlay of the two images.

results of any image registration performed between the reference and deformed images can be quantitatively compared to the *M*-DVF.

III.B. Validation of DIRs on prostate phantom

The fabricated prostate reference image was deformed using the *M*-DVF to generate a simulation image shown in Fig. 5(a). The deformable registrations were performed from the reference image to the simulation image using SR-Demons and MR-Demons with 500 iterations, respectively. The resultant DVFs were used to deform the reference image to the simulation image in the same manner that the simulation image was created. The images warped by the two Demons DVFs are shown in Figs. 5(b) and 5(c). In an analogous way, single resolution B-Spline registrations (using five and eight nodes) were performed from the reference to the simulation image with the resultant DVFs utilized to reconstruct the simulation images shown in Figs. 5(d) and 5(e).

Figure 5 shows results for the two Demons registrations. Large errors can be observed at the interface between the rectum and prostate [Figs. 5(b) and 5(c)]. Undesired deformation of the femoral heads is also noted [Figs. 5(d) and 5(e)]. Due to the uniform image intensities in the interior regions of the bladder, prostate, and femoral heads, the potential image registration errors at such locations will not be detected. However, using the *M*-DVF as a benchmark, potential registration errors can be identified on a voxel-by-voxel basis. The profiles in Fig. 6(a) illustrate that the SR-Demons registrations worked well at organ boundaries, but large discrepancies were observed within the interior of the prostate ($30 < Y < 80$). Also, these errors cannot be identified

visually, suggesting that landmark-based evaluation of DIR accuracy may potentially underestimate registration errors.

The MR approach may help reduce the registration errors in homogeneous regions (e.g., within the interiors of the bladder and prostate) if these regions undergo significant deformation. However, if the deformation is small this approach may produce increased errors [see the region of $120 < Y < 150$ in Fig. 6(a) for the rectum]. In general, the MR-Demons results are more accurate than those with a SR as shown in Table I.

Table I shows that the convergence of MR-Demons is faster than that of SR-Demons. However, even after 1600 iterations, these registrations still have large errors. This is suggestive that the accuracy of each registration is strongly influenced by the underlying image intensity gradients.

The results in Table II show that the multiresolution approach improves the accuracy of B-Spline-based registration. However, for the same resolution level, increasing the number of the grid nodes may result in reduced accuracy. For the SR-B-Spline registrations, results with five grid nodes agreed with FEM-based simulation within 2 mm. Increasing the number of nodes increased the resultant DVF errors in homogeneous regions [see Fig. 6(d)]. For the MR-B-Spline registrations, accuracy in the *y*-dimension was improved as the number of the nodes increased [Fig. 6(d)]; however, accuracy along the *z*-profiles was reduced [see Fig. 6(e)]. Note also that there are more image intensity variations in the *y*-direction than in the *z*-direction. This may suggest that the number of the grid nodes for an optimal B-Spline registration is affected by the intensity gradients of the underlying images.

FIG. 5. (a) The synthesized image reconstructed by the FEM model. The warped reference image by (b) SR-Demons, (c) MR-Demons, (d) SR-B-Spline with five nodes, and (e) SR-B-Spline with eight nodes, respectively.

FIG. 6. The profiles of the FEM benchmark and the displacements of (a) the Demons registrations on the Y axis (from anterior to posterior) shown in Fig. 5(a), (b) Demons at Z axis (from superior to inferior), (c) the Demons illustrated using the reverse comparison, (d) three SR-B-Splines at the Y axis, (e) three MR-B-Splines at Y axis, and (f) three MR-B-Splines at Z axis.

In the above discussion, the DVFs were compared to the M -DVF using the forward comparing method. The same registration algorithm may also be compared using a reverse transformation with the FEM benchmark (Sec II B). It can be seen that the DVFs shown in Fig. 6(c) are different from those in Fig. 6(a). This is mainly due to different edge effects in the simulation and reference images, as well round-off errors induced by the index conversion functions in the reverse comparison. To minimize these effects, regions deformed out of the boundary were excluded.

III.C. Validation of DIRs on lung patient CT images

SR-Demons and MR-Demons registrations were performed from the reference lung image [Fig. 4(a)] to the deformed lung [Fig. 4(b)]. Taken along the red line shown in Fig. 8(a), the z -profiles in Fig. 7(a) show that the SR-Demons with 200 iterations produces a large discrepancy in the vicinity of the diaphragm, as highlighted by the arrow in Fig. 8(c). Due to low image intensity the deformation error is large but is improved using the MR-Demons algorithm with 200 iterations [Figs. 8(b) and 8(d)]. A similar example is found in the

SR-Demons-based y -profile [Fig. 7(b)], corresponding to the line in Fig. 8(e), where the discrepancy produced by the SR-Demons registration is highlighted by the circles in Figs. 8(e) and 8(f). This error was minimized in the MR-Demons registration [Fig. 7(b)].

Since the tumor was modeled with different parameters (incompressible and less elastic) than the lung, the slope of the M -DVF profile is different in the tumor region [Fig. 7(b)]. The errors of the two Demons registrations in the region $150 < Y < 200$ are slightly larger than any other location outside of the tumor [Fig. 7(b)]. This may be attributed to the reduced image gradients within the tumor, similar to the case shown in Fig. 6(a). Table III shows the average DVF error of the SR and MR-Demons registrations. The MR approach converged faster than that using the SR method, especially in the lung.

The SR-B-Spline registration agreed with the FEM benchmarks in the lung [see $40 < Y < 150$ of Fig. 7(c)], with a large error noted in the region of the chest wall [Fig. 8(g)]. This discrepancy was reduced in the MR-B-Spline registration [Fig. 8(h)]. As shown in Table IV, the mean errors of the

TABLE I. Mean displacement errors of the Demons registrations computed on the prostate phantom.

Mean DVF error (mm)	Number of iterations					
	100	200	300	500	800	1600
SR-Demons	4.26	3.68	3.53	3.29	2.95	2.53
MR-Demons	2.40	2.36	2.31	2.23	2.13	2.01

TABLE II. Mean displacement errors of the B-Spline registrations computed on the prostate phantom.

DVF mean error (mm)	Number of nodes			
	5	8	12	16
SR-B-Spline	1.92	2.68	3.42	3.06
MR-B-Spline	1.57	1.61	2.09	1.85

B-Spline registrations averaged in the entire image domain are about 1.5 mm, but their errors averaged only in the lung region are as low as 0.5 mm.

The results in Table IV indicate that the accuracy of the SR-B-Spline registrations was improved when the number of the grid nodes was increased from $n=5$ to 10. The accuracy decreased as the n increased over 15 nodes. For the MR-B-Spline, the influence of the number of the grid nodes is negligible, though the computation time is affected dramatically. The accuracy of the SR-B-Spline or MR-B-Spline registrations does not appear to be proportional to the number of the grid nodes used. This is consistent with the observation that for images with low intensity gradients, e.g., the prostate case above, increasing the number of the grid nodes reduced the accuracy of both the SR-B-Spline and MR-B-Spline registrations.

In the development of the lung deformable model, the spine and ribs were kept rigid. Tables III and IV show that in all cases, mean errors averaged in the lung region are less than half of the errors averaged in the entire image domain. The results suggested that with the Demons and B-Spline registration algorithms, regions with different image gradients can be registered at different accuracies.

III.D. Detection of DIR errors

Registration errors can be identified by comparing target and warped images as shown in Fig. 8. However, if registration errors appear in regions of relatively uniform intensity (low contrast), these errors cannot be identified by the visual evaluation. For example, the two images in Figs. 9(a) and 9(b) show that the left diaphragm in the region of $75 < Z < 85$ was properly aligned in the SR-Demons registration. However, the profile of the displacement errors in Fig. 9(f) indicates large errors along the line of $X=300$ [illustrated in Fig. 9(a)]. The displacement error was calculated with respect to the FEM benchmark.

TABLE IV. Mean displacement errors of the B-Spline registrations computed on the lung CT images.

Mean DVF errors of B-Spline (mm)		The number of grid nodes		
		5	10	15
Averaged in image domain	SR-B-Spline	1.66	1.36	1.78
	MR-B-Spline	1.51	1.48	1.52
Averaged in the lung only	SR-B-Spline	0.72	0.50	0.50
	MR-B-Spline	0.61	0.46	0.47

The error profile in Fig. 9(f) shows large displacement errors in the intensity-uniform region ($75 < Z < 85$), where the image intensity comparison method [Figs. 9(a) and 9(b)] fails to identify these errors. The difficulty to detect such errors is due to lack of visible features in this area (low intensity gradient region), unlike in the lung region where variations in voxel intensities are present [Figs. 9(d) and 9(e)]. Moreover, in a clinical scenario, a benchmark model (as we developed here using FEM) to determine the displacement errors is not typically available. To address this issue, we previously proposed the concept of UE as a feasible method to detect displacement errors in the DIR process.³¹ To evaluate the association of UE with the DIR displacement errors, the DVF from the SR-Demons was substituted into the FEM framework to calculate its UE. The slice of the resultant UE image corresponding to Fig. 9(a) is demonstrated in Fig. 9(c), and the profiles of the displacement errors and the unbalanced energy along the line shown in Fig. 9(a) are compared in Fig. 9(f).

Figure 10(a) illustrates the displacement vector fields of the SR-Demons registration. The corresponding plot of unbalanced energy shown is Fig. 10(b). In the areas bounded by the ellipse and rectangle, the displacement vector field is associated with nonphysical deformation. These areas also correspond with the regions of high UE in Fig. 10(b). Although the UE values are not exactly correlated with the displacement errors, at this stage we propose UE as a feasible tool to determine which areas in clinical images will require more careful assessment, adjustment, or correction during image intensity based deformable registration. Further work related to the UE-based method to develop it into a more accurate tool for assessment of DIR displacement errors is in progress.

TABLE III. Mean displacement errors of Demons registrations computed on the lung CT images.

Mean of DVF errors (mm)		The number of iterations			
		100	200	300	500
Averaged in image domain	SR-Demons	2.34	1.63	1.40	1.33
	MR-Demons	1.76	1.39	1.35	1.30
Averaged in the lung only	SR-Demons	1.57	0.87	0.77	0.74
	MR-Demons	0.76	0.76	0.76	0.76

FIG. 7. (a) The Z displacement profiles of FEM, Demons, and B-Spline registrations on the line shown in Fig. 8(a) from superior to inferior. (b) and (c) The Y displacement profiles of Demons and B-Spline registrations on the line shown in Fig. 8(e) from anterior to posterior, respectively.

IV. DISCUSSION

Methods to properly validate DIR accuracy are being increasingly investigated, as image registration is a central requirement to accurate planning and delivery in IGRT. A recent, important investigation by Kashani *et al.*²⁶ evaluated eight different registration algorithms based on landmarks embedded within a physical phantom. While the use of physical phantoms provides an objective measure of the registration accuracy, one could argue that the presence of more uniform intensity regions within the phantom, as opposed to within clinical CT images (where more intensity gradients are present), may limit the applicability of phantom tests. In contrast, Castillo *et al.*²⁷ measured thousands of landmarks within patients' CT images and produced benchmarks for evaluation of DIR algorithms in clinical scenarios. In this study we chose to develop two FEMs, one of a low image gradient phantom, and the other of a lung patient CT data set. The generated FEM deformation field is independent of the similarity metric of the two registered images, and conse-

quently may serve as an alternative benchmark to investigate the effects of different parameters and image intensities on the performance of deformable image registrations.

The accuracy of each registration algorithm is largely determined by the parameters used for the registration. However optimal parameters for a given algorithm may vary significantly, depending on the image data sets being registered. For example, the MR-Demons registration with 100 iterations has converged to the minimum error (mean of 0.76 mm, Table III) in the lung where there are larger image intensity gradients. However, for the prostate case, which is relatively homogeneous (low intensity gradients), the registration after 1600 iterations still has a 2 mm mean error. For the B-Spline algorithms, the selection of the optimum number of nodes is sensitive to the intensity gradients. Increasing the number of nodes may decrease the accuracy of the registration for a structure with low intensity gradients (e.g., the prostate) as noted in Table II. Although we have explored optimal parameters for the registration algorithms proposed

FIG. 8. Top: The coronal slices of (a) the moving image, (b) the benchmark image synthesized using the FEM, and [(c) and (d)] the images warped by the displacements of the SR-Demons and the MR-Demons, respectively. Bottom: (e)–(h) Transverse slices of the images warped by the displacements of FEM, SR-Demons, SR-B-Spline, and MR-B-Spline, respectively.

FIG. 9. (a) The image warped by the FEM benchmark, (b) the image warped by the single resolution Demons, (c) the unbalanced energy of the single resolution Demons registration, (d) the FEM warped image with lung highlighted, (e) the Demons warped image with lung highlighted, and (f) the profiles of the demons displacement error and its unbalanced energy.

here, it is important to note that the algorithm accuracy is dependent on the implementation details. Variation in the accuracy of the results is possible even for a given algorithm depending on the implementation of specific procedures, such as the interpolation scheme. This point must be considered when interpreting the results of this study.

For a patient lung image data set, separate evaluations of the registrations were performed, considering the entire image, as well as the lung only. The results show that the displacement errors were reduced more than 50% when the registrations were evaluated in the lung only (see Tables III and IV), where the presence of naturally occurring structures of varying intensities (producing larger intensity gradients) helped improve the registration accuracy. This may suggest that landmark or feature-based evaluation methods may underestimate the overall registration errors even when the landmarks are registered accurately because the errors in regions of uniform intensities are not properly accounted for. The additional information contained in the unbalanced elas-

ticity energies (Fig. 9) may help identify the requirement for a more detailed analysis of these areas in clinical image registrations.

Recently, Werner *et al.*³⁰ showed that the FEM modeling of lung deformation will have large errors if the properties of the tumor are not included in the model. In this study, the developed FEM phantoms were designed for the entire image domain with different elasticity and compressibility parameters assigned to different organs based on their image intensities. This approach is able to accommodate the presence of tumor vs lung or rigid bones vs soft tissue as well volume changes in bladder, rectum, and lung. The combination of mechanical models with clinical CT images renders the simulated deformations more realistic, thereby providing a platform to investigate various issues in deformable image registration.

In this study, the simulation image was reconstructed from a reference image using the lung model configured with heuristic forces applied to the diaphragm for illustration purposes. Many other important factors affecting registration accuracy, such as CBCT artifacts, tumor shrinkage, CT density variations, and conditions of large deformation are not accounted, so the resultant simulation results may not be reflective of clinical circumstances, and further improvement of the simulation reality is required before any clinical use.

V. CONCLUSION

In this study, two computational phantoms have been developed for analysis of the accuracy of the Demons and B-Spline registrations using different parameters and images with varying intensity gradients. Results show that registra-

FIG. 10. (a) The displacement vectors of the SR-Demons registration; (b) its corresponding unbalanced energy map.

tion errors, which are often caused by calculations which have not converged or by inappropriately chosen parameters in the algorithms, are prone to appear in regions with uniform image intensity (low intensity gradient regions). Special attention must be paid to these regions in the clinic applications of these DIR algorithms.

- ^{a)} Author to whom correspondence should be addressed. Electronic mail: hzhong1@hfh.org; Telephone: (313) 916-8786.
- ¹M. L. Kessler, "Image registration and data fusion in radiation therapy," *Br. J. Radiol.* **79**, S99–S108 (2006).
 - ²D. Yan, "Developing quality assurance processes for image-guided adaptive radiation therapy," *Int. J. Radiat. Oncol., Biol., Phys.* **71**, S28–S32 (2008).
 - ³M. Chao, Y. Xie, and L. Xing, "Auto-propagation of contours for adaptive prostate radiation therapy," *Phys. Med. Biol.* **53**, 4533–4542 (2008).
 - ⁴D. Sarrut, "Deformable registration for image-guided radiation therapy," *Z. Med. Phys.* **16**, 285–297 (2006).
 - ⁵Y. Yang, E. Schreibmann, T. Li, C. Wang, and L. Xiang, "Evaluation of on-board kV cone beam CT (CBCT)-based dose calculation," *Phys. Med. Biol.* **52**, 685–705 (2007).
 - ⁶D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," *Phys. Med. Biol.* **42**, 123–132 (1997).
 - ⁷W. Lu, G. H. Olivera, Q. Chen, K. J. Ruchala, J. Haimerl, S. L. Meeks, K. M. Langen, and P. A. Kupelian, "Deformable registration of the planning image (kVCT) and the daily images (MVCT) for adaptive radiation therapy," *Phys. Med. Biol.* **51**, 4357–4374 (2006).
 - ⁸M. Rosu, J. M. Balter, I. J. Chetty, M. L. Kessler, D. L. McShan, P. Balter, and R. K. Ten Haken, "How extensive of a 4D data set is needed to estimate cumulative dose distribution plan evaluation metrics in conformal lung therapy?," *Med. Phys.* **34**, 233–245 (2007).
 - ⁹W. Lu, M. L. Chen, G. H. Olivera, K. J. Ruchala, and T. R. Mackie, "Fast free-form deformable registration via calculus of variations," *Phys. Med. Biol.* **49**, 3067–3087 (2004).
 - ¹⁰T. Zhang, Y. Chi, E. Meldolesi, and D. Yan, "Automatic delineation of on-line head-and-neck computed tomography images: Toward on-line adaptive radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **68**, 522–530 (2007).
 - ¹¹M. R. Kaus, K. K. Brock, V. Pekar, L. A. Dawson, A. M. Nichol, and D. A. Jaffray, "Assessment of a model-based deformable image registration approach for radiation therapy planning," *Int. J. Radiat. Oncol., Biol., Phys.* **68**, 572–580 (2007).
 - ¹²K. K. Brock, M. B. Sharpe, L. A. Dawson, S. M. Kim, and D. A. Jaffray, "Accuracy of finite element model-based multi-organ deformable image registration," *Med. Phys.* **32**, 1647–1659 (2005).
 - ¹³G. E. Christensen and H. J. Johnson, "Consistent image registration," *IEEE Trans. Med. Imaging* **20**, 568–582 (2001).
 - ¹⁴A. Leow, S. C. Huang, A. Geng, J. Becker, S. Davis, A. Toga, and P. Thompson, "Information Processing in Medical Imaging," Lecture Notes in Computer Science Vol. 3565 (2005).
 - ¹⁵D. Yang, H. Li, D. A. Low, J. O. Deasy, and I. E. Naqa, "A fast inverse consistent deformable image registration method based on symmetric optical flow computation," *Phys. Med. Biol.* **53**, 6143–6165 (2008).
 - ¹⁶E. T. Bender and W. A. Tome, "The utilization of consistency metrics for error analysis in deformable image registration," *Phys. Med. Biol.* **54**, 5561–5577 (2009).
 - ¹⁷H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," *Phys. Med. Biol.* **50**, 2887–2905 (2005).
 - ¹⁸T. Guerrero, G. Zhang, T. C. Huang, and K. P. Lin, "Intrathoracic tumour motion estimation from CT imaging using the 3D optical flow method," *Phys. Med. Biol.* **49**, 4147–4161 (2004).
 - ¹⁹W. R. Crum, C. Tanner, and D. J. Hawkes, "Anisotropic multi-scale fluid registration: Evaluation in magnetic resonance breast imaging," *Phys. Med. Biol.* **50**, 5153–5174 (2005).
 - ²⁰M. Tashiro, S. Minohara, T. Kanai, K. Yusa, H. Sakurai, and T. Nakano, "Three-dimensional velocity mapping of lung motion using vessel bifurcation pattern matching," *Med. Phys.* **33**, 1747–1757 (2006).
 - ²¹M. Söhn, M. Birkner, Y. Chi, J. Wang, D. Yan, B. Berger, and M. Alber, "Model-independent, multimodality deformable image registration by local matching of anatomical features and minimization of elastic energy," *Med. Phys.* **35**, 866–878 (2008).
 - ²²E. Heath, J. Seco, Z. Wu, G. C. Sharp, H. Paganetti, and J. Seuntjens, "A comparison of dose warping methods for 4D Monte Carlo dose calculation in lung," *J. Phys.: Conf. Ser.* **102**, 1–7 (2008).
 - ²³D. Sarrut, V. Boldea, S. Miguet, and C. Ginestet, "Simulation of 4D CT images from deformable registration between inhale and exhale breath-hold CT scans," *Med. Phys.* **33**, 605–617 (2006).
 - ²⁴M. Serban, E. Heath, G. Stroiian, D. L. Collins, and J. Seuntjens, "A deformable phantom for 4D radiotherapy verification: Design and image registration evaluation," *Med. Phys.* **35**, 1094–1102 (2008).
 - ²⁵R. Kashani, M. Hub, M. L. Kessler, and J. M. Balter, "Technical note: A physical phantom for assessment of accuracy of deformable alignment algorithms," *Med. Phys.* **34**, 2785–2788 (2007).
 - ²⁶R. Kashani, M. Hub, J. M. Balter, M. L. Kessler, L. Dong, L. Zhang, L. Xing, Y. Xie, D. Hawkes, J. A. Schnabel, J. McClelland, S. Joshi, Q. Chen, and W. Lu, "Objective assessment of deformable image registration in radiotherapy: A multi-institution study," *Med. Phys.* **35**, 5944–5953 (2008).
 - ²⁷R. Castillo, E. Castillo, R. Guerra, V. E. Johnson, T. McPhail, A. K. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.* **54**, 1849–1870 (2009).
 - ²⁸T. Zhang, N. Orton, T. Mackie, and B. Paliwal, "Technical note: A novel boundary condition using contact elements for finite element based deformable image registration," *Med. Phys.* **31**, 2412–2415 (2004).
 - ²⁹A. Al-Mayah, J. Moseley, and K. K. Brock, "Contact surface and material nonlinearity modeling of human lungs," *Phys. Med. Biol.* **53**(1), 305–317 (2008).
 - ³⁰R. Werner, J. Ehrhardt, R. Schmidt, and H. Handels, "Patient-specific finite element modeling of respiratory lung motion using 4D CT image data," *Med. Phys.* **36**, 1500–1511 (2009).
 - ³¹H. Zhong, T. Peters, and J. V. Siebers, "FEM-based evaluation of deformable image registration for radiation therapy," *Phys. Med. Biol.* **52**, 4721–4738 (2007).
 - ³²H. Zhong, M. P. Wachowiak, and T. M. Peters, "A real time finite element based tissue simulation method incorporating nonlinear elastic behavior," *Comput. Methods Biomech. Biomed. Eng.* **8**, 177–189 (2005).
 - ³³R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods* (SIAM, Philadelphia, 1994).
 - ³⁴O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method: Volume 1, The Basis*, 5th ed. (Butterworth Heinemann, Barcelona, Spain, 2000).
 - ³⁵H. Zhong and J. Siebers, "Monte Carlo dose mapping on deforming anatomy," *Phys. Med. Biol.* **54**, 5815–5830 (2009).
 - ³⁶J. P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Image Anal.* **2**, 243–260 (1998).
 - ³⁷L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The itk Software Guide*. 2nd Edition, Kitware, Inc., 2005.
 - ³⁸D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imaging* **18**(8), 712–721 (1999).