Characterization of deformation and physical force in uniform low contrast anatomy and its impact on accuracy of deformable image registration

Raj Varadhan^{a)}

*Department of Radiation Oncology, University of Minnesota, Minneapolis, Minnesota 55455
and Minneapolis Radiation Oncology, Minneapolis, Minnesota 55432*

Taiki Magome and Susanta Hui

Department of Radiation Oncology, Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455

(Received 4 July 2015; revised 11 November 2015; accepted for publication 16 November 2015; published 17 December 2015)

Purpose: Little is known about the effect of force on organ deformation and consequently its impact on precision dose delivery. The purpose of this study was to evaluate the fundamental relationship between anatomic deformation and its causative physical force to ascertain if a threshold limit exists for deformable image registration (DIR) accuracy in uniform low contrast anatomy, beyond which its applicability may be clinically inappropriate.

Methods: To simulate a simplified model, a tissue equivalent deformable bladder phantom with 21 implanted fiducial markers was developed using a viscoelastic polymer. The bladder phantom was deformed by applying a force in increments from 10 to 70 N. DIR accuracy was studied using intensity based μM and Velocity B-spline algorithms by comparing the 3D vector of the 21 marker locations at the original target image with the synthetically derived marker positions from each target image obtained from DIR.

Results: The relationship between applied force in 1D deformation along the axis of applied force and 3D deformation of the phantom showed a linear response. The maximum and average displacements of markers exhibited a nonlinear response to the applied force. In the absence of implanted markers, DIR performance was suboptimal with a threshold limit of only 20 N (5 mm deformation) beyond which the average marker error was ≥ 3 mm. DIR performance improved significantly with the addition of only one marker for the intensity based μM algorithm. In contrast, the Velocity B-spline algorithm showed reduced sensitivity to the number of markers introduced in both the source and target images.

Conclusions: The limits of applicability of DIR are strongly dependent on the magnitude of deformation. There is a threshold limit beyond which the accuracy of DIR fails in uniform low contrast anatomy. The sensitivity of the DIR performance to the number of fiducial markers present indicates that if DIR performance is solely assessed with the contrast rich features present in clinical anatomy, the results may not be reflective of the true DIR performance in uniform low contrast anatomy.

© 2016 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4937935>]

Key words: deformable image registration, organ deformation, uniform low contrast anatomy, deformable phantom

1. INTRODUCTION

Deformable image registration (DIR) has gained widespread acceptance with the availability of several commercial DIR platforms. It is well known that the accuracy of DIR varies depending on the algorithm used and that the suitability of a particular algorithm may be site specific.^{1,2} Previous studies have evaluated DIR accuracy through physical phantoms or synthetically induced deformation on patient CT study sets.^{3–8} These studies relied on either the identification of high contrast features, like bony anatomy, or through implanted fiducial markers for DIR validation. However, little is known about the accuracy of DIR in uniform low contrast anatomy (e.g., liver, pancreas, kidney, prostate, stomach, bladder, and bowel⁹) where the DIR calculation is driven primarily by contrast differences at the region boundary. The accuracy of DIR in low

contrast regions is particularly important for the commercially available algorithms that do not allow parameter optimization. It is known that different operational parameters, such as smoothness criteria, can greatly influence the registration accuracy.^{10–12} The validation of DIR accuracy in uniform low contrast anatomy is also of interest in the context of dose warping for the purposes of dose accumulation.^{13,14}

Although DIR accuracy has been studied previously in the context of uniform low contrast regions, it was mostly used to describe arbitrary deformations using a deformable gel.⁹ The fundamental relationship between different levels of deformation due to an applied force and the existence of a threshold limit for DIR in uniform low contrast anatomy have not been fully understood.

Soft tissue organ deformation in the context of a medical simulator has been previously modeled as a biofeedback

between physics (force, deformation, density, kinematics, etc.) and the anatomy and physiology of organs.^{15–19} In a radiation therapy context, the occurrence of a tumor in an organ (physiological process) will modify the biomechanical properties of the tissue (physics). For example, it is known that the mean density of parotid glands (physics) varies during the course of head and neck intensity modulated radiation therapy (IMRT) therapy due to changes in tumor size and salivary flow (physiology).²⁰ A quantitative relationship between force and deformation may provide insights into the deformation characteristics of other organs with known biomechanical properties and has the potential to create simulated organ deformations. A force–deformation relationship of organs has important applications in real time surgical simulations that incorporate haptic feedback, optimization of surgical tool design, creating “smart” instruments capable of assessing pathology or force-limiting novice surgeons, and understanding the mechanism of tissue injury and damage thresholds.^{17,21–24}

The purpose of this study is to (a) characterize the relationship between deformation of an organ and the applied physical force, (b) ascertain and establish a threshold deformation limit in uniform low contrast anatomy for accuracy of commercial DIR algorithms, and (c) evaluate if a minimum number of fiducial markers are required in uniform low contrast anatomy for a threshold 3 mm DIR accuracy.

2. METHODS AND MATERIALS

2.A. Deformable bladder phantom and deformation

A deformable phantom mimicking human “bladderlike” organ volume and geometry was made from a solid water prototype. The phantom has dimensions of $10.5 \times 9 \times 4.8$ cm measuring 386 cm^3 in total volume. The prototype was coated with mold release and a silicone rubber compound was poured around it. It was then placed in a vacuum chamber at a pressure of 635 mm of mercury for approximately 20 h to create a mold. The mold was filled with Akton viscoelastic polymer (Action Products, Hagerstown, MD) and the deformable phantom was removed from the mold after hardening. The Akton viscoelastic polymer used is tissue equivalent with a relative electron density of 1.02 and physical density of 1.03 g/cm^3 .²⁵ It has a peak tensile strength of 157 kPa and Young’s modulus of 17.9 kPa.²⁶ The mechanical and tensile properties of the phantom are comparable to the human bladder which has a Young’s modulus of 16 kPa,^{27,28} density of 1.04 g/cm^3 ,²⁹ and peak tensile strength of 270 ± 140 kPa.³⁰ An air-dry synthetic rubber which resists moisture and absorption was applied on the phantom. Twenty one aluminum markers (18 markers with a diameter of 1.5 mm and 3 markers with a diameter of 2.5 mm) were implanted at depth and distributed at various axial planes in the phantom. The larger markers of 2.5 mm diameter were implanted in the superior, central, and inferior regions of the bladder phantom. A coronal 3D reconstructed view of the 21 markers is shown in Fig. 1(a). The bladder phantom was placed into an apparatus for measuring force–deformation properties [Fig. 1(b)]. The

acrylic apparatus had a mechanical piston at one end fitted with a customized miniature load cell with National Institute of Standards and Technology (NIST) traceable calibration. The load cell was connected to a digital process meter and controller which displayed the applied force. A precision weight was used to calibrate the load cell and readout. This setup demonstrated excellent reproducibility and linearity of the applied force induced by the piston on the phantom.³¹ The applied force on the phantom was varied incrementally from 10 to 70 N along the longitudinal axis (+z axis) of the phantom causing deformation in all three dimensions. The +z axis was chosen for the direction of applied force since the markers were uniformly distributed along the longitudinal axis and it also corresponded to the direction in which maximum compression could be applied to the phantom. Deformation was quantified both in terms of 1D and 3D deformation. 1D deformation was quantified using the maximum compression observed along the direction of applied force (+z axis). 3D deformation was quantified by the 95th percentile Hausdorff distance (95% HD).³² The 95% HD measures the maximum of the closest distance between two volumes, computed for each vertex of the two volumes. To calculate the 95% HD, the external contour of the deformable phantom in each deformation state was compared against the external contour in the original (undeformed) state using Slicer RT.³³ Using the 95% HD ensured that the outliers were rejected.

2.B. Deformable image registration algorithms

In this work, the accuracy of commercially available DIR algorithms was assessed, namely, the free-form intensity based deformation from MIM 6.3 (MIM Software, OH) and the modified B-spline algorithm in both the single pass and multipass modes from Velocity AI 3.0 (Velocity Medical Solutions, GA). MIM and Velocity were chosen due to the contrasting nature of these two algorithms. MIM uses a free form, fully automatic, intensity based deformation with a multiresolution approach which seeks to minimize intensity differences between two images.³⁴ MIM employs sum of squared differences (SSDs) between volumes as the similarity metric using a gradient descent optimizer. MIM also allows the user to manually inspect (Reg Reveal™) and refine (Reg Refine™) the deformation. Reg Reveal displays the original nondeformed secondary image rigidly aligned to the primary image, while still representing the underlying deformation matrix by approximating the local deformation field with the least squares, best fit rigid registration. In essence, Reg Refine allows the user to bias the DIR to achieve results similar to an adjusted local rigid registration.

Velocity AI uses a modified B-spline algorithm³⁵ combined with Mattes mutual information metric.³⁶ The number of control points in Velocity AI can be varied depending on the choice of multipass or single pass registration modes in the software. In the single pass mode, the finest grid resolution is applied. In the multipass mode, the grid resolution spacing begins with the coarse setting and gradually decreases to the finest resolution, potentially decreasing the convergence time of the optimizer. Since the degrees of freedom changes

FIG. 1. (a) Coronal view of the 21 implanted markers in the bladder phantom. (b) Apparatus for investigating force–deformation properties of the bladder phantom.

with volume of interest (VOI) in Velocity AI, in order to keep the VOI constant in all analyses, it was fixed as a circumscribed cuboid around the original target image phantom. The multipass algorithm is the default registration mode recommended by the vendor for clinical applications. In both the DIR platforms, the details of the algorithm parameters, including smoothness (regularization) criteria, are proprietary and are not user-defined.

2.C. DIR accuracy validation

The accuracy of DIR was validated in this study using the methodology outlined in the flow chart (Fig. 2). The deformable phantom was placed inside the force–deformation apparatus. CT images with $0.289 \times 0.289 \times 1$ mm voxel thickness were obtained in the original and multiple deformed geometries. It should be noted that the implanted markers did not bias the DIR calculation algorithms as they were synthetically masked from the source and target images by editing the Hounsfield units (HU) of the markers to equal the average HU of the phantom (10 HU). The marker positions were known explicitly by direct CT measurement. The 21 markers were individually contoured on each axial slice of the target image and each of the source images at various deformation states (10–70 N). Using the marker contours, the centroid of each marker was computed in spatial coordinates (x , y , z). After DIR was performed, the deformation vector field (DVF) was applied to the marker contours of the deformed source image. The centroid (x , y , z) of the warped target image contours from each source image (10–70 N) was then compared to the ground truth centroid coordinates of the original target image to calculate the 3D error. Further, the minimum number of fiducial markers necessary to arrive at a threshold 3 mm DIR accuracy within the uniform low contrast anatomy was investigated. This was accomplished by “unmasking” a linear sequence of corresponding markers (1, 3, 5, 7, 9, 13, 17, and 21 markers) in both source and target images. The 2.5 mm diameter marker at the center of the bladder was the first marker unmasked. Subsequent markers introduced in the images were distributed uniformly across the bladder phantom as shown in Fig. 3. The DVF obtained from DIR was applied to the source image containing all 21 markers to synthetically derive the target position of each individual marker.

The images were processed and cropped to remove the acrylic plates holding the phantom. The markers were synthetically removed from the target image before

performing DIR. After the DIR was performed, the DVF was applied to the raw source image containing the markers instead of the source image with markers synthetically masked. DIR accuracy was measured by comparing the 3D vector of the 21 implanted markers at the original CT (ground truth) with the synthetically derived marker positions from each target image obtained from DIR.

3. RESULTS

3.A. Response to applied force vs deformation

The applied force on the deformable phantom was incrementally varied from 10 to 70 N along the $+z$ axis, deforming the phantom as shown in Fig. 4. This mass and density preserving deformation ranged from 3 to 28 mm along z -axis, representing a bilateral compression of equal displacement from both sides. The deformation showed a linear response to the applied force with $R^2 = 0.99$. The 3D deformation as quantified by the 95% HD varied from 2 to 16 mm with $R^2 = 0.98$ as shown in Fig. 5(a). Figure 5(b) shows the Dice similarity coefficient (DSC) which estimates the spatial overlap between the original undeformed organ external surface and each deformed organ as a function of applied force. DSC represents the extent of external shape changes in the organ. As seen in Fig. 5(b), the DSC was ≥ 0.9 between 10 and 30 N of applied force indicating that external shape changes of the organ, compared to the organ in original state, were minimal up to 30 N. The average and maximum displacements of the 21 markers varied exponentially with $R^2 = 0.98$ and were in a direction orthogonal to applied force as shown in Fig. 5(c).

3.B. 3D vector difference between markers

3.B.1. All 21 markers present in source and target images

The average error was calculated from the 3D vector between each measured marker location and its synthetically calculated position. When all the high contrast markers were present in both the source and target images, the DIR accuracy was excellent (≤ 2 mm) across the entire deformation range for mim. The Velocity multipass algorithm also performed well (≤ 3 mm error), except at 70 N, while the Velocity single pass algorithm showed the largest spatial error as shown in Fig. 6(a).

FIG. 2. DIR validation workflow used in this study.

FIG. 3. Coronal view illustrating the spatial location of markers progressively introduced in both source and target images to assess if a minimum number of markers were required to arrive at an optimal DIR accuracy in uniform low contrast anatomy.

3.B.2. All 21 markers synthetically masked from source and target images

When all the markers were synthetically masked from both the source and target images, the DIR performance was suboptimal (≥ 3 mm error) beyond a force of only 20 N (5 mm deformation) as shown in Fig. 6(b). At deformation forces larger than 50 N, the Velocity multipass

algorithm performed better than default intensity based MIM algorithm. Using the Reg Refine™ tool in MIM, the average marker error decreased from 10.6 to 3.8 mm at the largest deformation (70 N) studied. In all the algorithms, the maximum error was along the y -axis, orthogonal to the axis of applied force, corresponding to the direction of maximum displacement of markers (y -axis) for a given applied deformation.

FIG. 4. Various deformation states of the phantom in response to an applied force from 10 to 70 N.

3.B.3. Sensitivity of DIR performance to number of fiducial markers

The influence of the number of fiducial markers on DIR accuracy was investigated at 30 N (13 mm deformation) and

70 N deformations (28 mm deformation) and shown in Figs. 7(a) and 7(b), respectively. DIR accuracy was excellent for MM (<2 mm) even with the addition of only 1 marker (central 2.5 mm marker) across the entire deformation range. There were no further improvements in DIR accuracy for

FIG. 5. (a) Relationship between applied force and deformation (1D and 3D) observed in the deformable phantom. (b) DSC comparing spatial overlap between the external organ surface of the original target image and each deformed state of the organ as a function of applied force. (c) Average and maximum 3D vector marker displacement as a function of applied force.

FIG. 6. (a) Comparison of the average displacement error for the deformation range studied when all the markers were present. (b) Comparison of average displacement error for the deformation range studied when all the markers were synthetically masked in the source and target images.

MIM when additional fiducial markers were introduced. For the Velocity multipass algorithm, while the average error decreased with a minimum of 3 markers at 30 N, similar improvements were not seen at 70 N. The B-spline algorithms used in Velocity AI showed reduced sensitivity to the number of fiducial markers present in the source and target images when compared to the intensity based MIM algorithm.

4. DISCUSSION

We have established an upper threshold limit of 20 N (5 mm 3D deformation) for MIM and Velocity multipass algorithms (≤ 3 mm error) in uniform low contrast anatomy [Fig. 6(b)]. The 3 mm error was chosen as a threshold for DIR accuracy because it has been shown that, when compared to MIM, the Velocity multipass B-spline algorithm results in a larger fraction of voxels showing more than 2 mm error. This is due to the smoothing of images at multiple resolutions at the initial registration which results in a residual error (~ 1 –2 mm) during image interpolation between voxels.² The linear response observed in this study between applied force and deformation is typical of the biomechanical properties of the human bladder, in contrast to other pelvic organs like the rectum, vaginal tissue, and liver, which show increased resistance to deformation with applied force.^{37,38}

Table I summarizes the average (Δ_{Avg}), the standard deviation (SD), and the maximum (Δ_{Max}) marker error for various deformation states when the 21 markers were present and masked from both the source and target images. The maximum marker error for all algorithms was approximately 1.5–3 times the average error.

The Reg Refine algorithm reduced errors when the markers were masked due to the fact that the known bilateral symmetric compression of the phantom created easily identifiable common points of interest on the surface, around the point of flexion. However, such improvement may not be possible in clinical patient geometry because of the inability to accurately track an organ's boundary for registration of the common points of interest.

It is interesting to note that even the addition of a single fiducial marker remarkably improved the DIR performance in MIM, while such improvement was not seen for Velocity AI. This was likely because the B-spline algorithm used in Velocity AI defines the deformation only on a sparse lattice of nodes overlaid on the image, and the displacement at any voxel is obtained by interpolation from the closest lattice nodes. Hence, its sensitivity to markers was potentially less than the intensity based matching used in MIM. Also interestingly, the spatial location of the single fiducial marker introduced in the image (superior, central, or the inferior 2.5 mm marker)

FIG. 7. (a) Comparison of the average displacement error for the deformation range studied as a function of the number of markers at 30 N deformation. (b) Comparison of the average displacement error for the deformation range studied as a function of number of markers at 70 N deformation.

did not have an impact on the DIR performance in MIM (data not shown). It is also not surprising that the single pass registration mode in Velocity AI performed poorly as it has been previously shown that a larger number of iterations are required at coarser resolution before resampling at a finer resolution to achieve optimal DIR accuracy.³⁹

Validation of DIR in the internal anatomy of organs in the pelvic region is often challenging because of the lack of visible landmarks. As shown in this study, although boundary matching appeared perfect even at 70 N (28 mm deformation), this did not guarantee the accuracy inside the organ of interest. It is known that the optimization of parameters used in DIR algorithms has a significant impact on low contrast DIR accuracy.^{9,40} MIM uses an intensity based algorithm which seeks to minimize the intensity differences between two images, while the B-spline algorithm used by Velocity AI tries to balance both the intensity information regularized by the inherent cost function and the spatial information regularized by the smoothness criteria.² The presence of uniform low contrast regions throughout the deformable bladder phantom made it particularly challenging for both the commercial algorithms studied because of the lack of intensity differences in the internal anatomy of the phantom. In the absence of user ability to edit the deformation parameters in the commercial DIR platforms, both the algorithms interpolated the deformation incorrectly in uniform low contrast regions

with deformation force beyond 20 N causing the resultant errors. The suboptimal spatial accuracy of DIR in uniform low contrast anatomy indicates that the dose warping employed by commercial DIR platforms remains questionable especially at larger deformations.

Prior studies to verify accuracy of DIR in uniform low contrast anatomy have not established either a threshold deformation or a minimum number of fiducial markers necessary to arrive at clinically acceptable DIR spatial accuracy. To our knowledge, this is the first study in the context of DIR validation where deformation was parameterized with its causative force and a threshold limit derived for the accuracy of commercial DIR algorithms. Yeo *et al.*⁹ investigated DIR accuracy of 12 algorithms, for 3 geometric deformations, using markers implanted in a deformable gel, and concluded that the original Horn and Schunck algorithm⁴¹ yielded the best results. They found poor agreement for the spatial accuracy of marker locations using the commercial Velocity B-spline algorithm. Wognum *et al.*⁴² used a porcine bladder with markers implanted at the surface of the bladder and evaluated the DIR accuracy by varying bladder volume from 100 (fixed image) to 400 ml. They concluded that there was a significant increase in marker error with increasing bladder volume for all algorithms studied with the exception of the commercial Velocity B-spline algorithm. However, accuracy of markers on the surface does not guarantee

TABLE I. The average, standard deviation, and maximum errors between the true and synthetically derived marker positions.

Magnitude of 3D marker error							
Deformation force (N)	Algorithm	All 21 markers present			All 21 markers masked		
		Δ_{Avg} (mm)	SD (mm)	Δ_{Max} (mm)	Δ_{Avg} (mm)	SD (mm)	Δ_{Max} (mm)
10	MIM	1.4	1.3	4.5	1.4	0.86	3.3
	Velocity multipass	0.8	0.4	1.6	2.2	0.8	3.9
	Velocity single pass	3.3	1.0	5.5	5.7	1.3	9.1
20	MIM	1.0	1.0	3.6	2.0	1.5	5.4
	Velocity multipass	1.7	1.4	4.9	2.4	0.8	4.7
	Velocity single pass	3.9	1.5	7.4	5.7	1.9	8.4
30	MIM	1.3	1.0	3.2	4.3	2.0	8.3
	MIM Reg Refine				1.5	0.8	3.5
	Velocity multipass	1.5	1.5	5.0	4.1	1.7	6.6
	Velocity single pass	2.8	1.0	5.2	6.1	1.9	8.7
40	MIM	1.4	0.9	3.8	5.1	1.7	9.0
	MIM Reg Refine				2.3	1.3	4.8
	Velocity multipass	2.4	1.7	5.2	4.1	1.7	6.8
	Velocity single pass	2.8	1.2	5.0	5.7	3.9	14.5
50	MIM	1.6	1.0	4.2	5.3	2.8	8.8
	MIM Reg Refine				3.6	0.8	5.2
	Velocity multipass	2.5	1.5	6.0	3.9	1.6	6.7
	Velocity single pass	4.4	1.2	6.9	5.9	1.7	9.3
70	MIM	1.9	1.5	5.2	10.6	3.8	16.4
	MIM Reg Refine				3.8	2.1	8.6
	Velocity multipass	4.0	2.3	9.2	4.9	2.5	8.7
	Velocity single pass	5.8	1.7	8.8	7.5	2.1	11.0

the spatial accuracy inside the organ of interest. Similar to the findings of Yeo *et al.*⁹ in our study, we found that even though boundary matching appeared correct even at 70 N, this did not guarantee accuracy of markers implanted inside the organ. Very recently, Velec *et al.*⁴³ showed that the accuracy of DIR in uniform low contrast anatomy can be improved over intensity based algorithms by using biomechanical DIR. Biomechanical DIR algorithms that use finite element analysis to describe boundary conditions and allow fitting linear elastic material properties like Young's modulus and Poisson's ratio to match the biomechanical properties of the deforming organ would provide better accuracy of DIR in uniform low contrast anatomy. A quantitative force–deformation relationship should better enable biomechanical DIR algorithms to derive optimal DIR accuracy.

The deformable phantom used in this study had a uniform CT number (± 10 HU) and density similar to organs such as the bladder, prostate, pancreas, stomach, kidney, liver, breast, and diaphragm in human anatomy and the results in this study would apply to those organs. The DIR results do not apply to density changing anatomy like the lung or where the mass is not conserved (full vs empty bladder, organ atrophy, tumor inflammation, etc.). One of the limitations of the study was that although bladder geometry was chosen for the shape

of the phantom, the applied deformation especially at larger deformations may not represent clinically observed changes in the bladder which are typically due to emptying and filling of organ.⁴² Although symmetric bilateral compression was studied, future study will include asymmetric compression, changes in the direction of applied force, and 3D compression. The viscoelastic polymer used can be molded to any organ shape and has the potential to adjust its tensile properties to match other organs in human anatomy which will be the subject of future work.

5. CONCLUSION

A force–deformation relationship was quantified for a tissue equivalent deformable bladder phantom. The limit of applicability of DIR was strongly dependent on the magnitude of deformation. Validation of spatial accuracy using markers for commercially available DIR algorithms in uniform low contrast anatomy revealed a threshold limit of only 5 mm deformation. For the intensity based MIM algorithm, the addition of even one high contrast fiducial marker greatly improved the DIR accuracy. The sensitivity of the DIR performance to the number of fiducial markers present indicates that if the DIR performance was solely assessed with the various contrast rich features present in

clinical anatomy, the results may not be reflective of the true DIR performance in uniform low contrast anatomy.

ACKNOWLEDGMENT

This work was supported by National Institute of Health Grant No. 1R01CA154491.

^{a)}Author to whom correspondence should be addressed. Electronic mail: raj@mropa.com

¹N. Kirby, C. Chuang, U. Ueda, and J. Pouliot, "The need for application-based adaptation of deformable image registration," *Med. Phys.* **40**(1), 011702 (10pp.) (2013).

²K. Nie, C. Chuang, N. Kirby, S. Braunstein, and J. Pouliot, "Site-specific deformable imaging registration algorithm selection using patient-based simulated deformations," *Med. Phys.* **40**(4), 041911 (10pp.) (2013).

³K. K. Brock, "Results of a multi-institution deformable registration accuracy study (MIDRAS)," *Int. J. Radiat. Oncol., Biol., Phys.* **76**(2), 583–596 (2009).

⁴R. 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**(7), 1849–1870 (2009).

⁵R. Kashani *et al.*, "Objective assessment of deformable image registration in radiotherapy: A multi-institution study," *Med. Phys.* **35**(12), 5944–5953 (2008).

⁶N. Kirby, C. Chuang, and J. Pouliot, "A two-dimensional deformable phantom for quantitatively verifying deformation algorithms," *Med. Phys.* **38**(8), 4583–4586 (2011).

⁷R. Varadhan, G. Karangelis, K. Krishnan, and S. Hui, "A framework for deformable image registration validation in radiotherapy clinical applications," *J. Appl. Clin. Med. Phys.* **14**(1), 192–213 (2013).

⁸H. Zhong, J. Kim, and I. J. Chetty, "Analysis of deformable image registration accuracy using computational modeling," *Med. Phys.* **37**(3), 970–979 (2010).

⁹U. J. Yeo, J. R. Supple, M. L. Taylor, R. Smith, T. Kron, and R. D. Franich, "Performance of 12 DIR algorithms in low-contrast regions for mass and density conserving deformation," *Med. Phys.* **40**(10), 101701 (12pp.) (2013).

¹⁰M. L. Taylor, U. J. Yeo, T. Kron, J. Supple, S. Siva, D. Pham, and R. D. Franich, "Comment on 'It is not appropriate to 'deform' dose along with deformable image registration in adaptive radiotherapy' [Med. Phys. **39**, 6531–6533 (2012)]," *Med. Phys.* **40**(1), 017101 (3pp.) (2013).

¹¹W. O. G. Lu *et al.*, "Deformable registration of the planning image (kvCT) and the daily images (MVCT) for adaptive radiation therapy," *Phys. Med. Biol.* **51**(17), 4357–4374 (2006).

¹²G. J. Klein and R. H. Huesman, "Elastic material model mismatch effects in deformable motion estimation," *IEEE Trans. Nucl. Sci.* **47**(3), 1000–1005 (2000).

¹³M. Velec, J. L. Moseley, T. Craig, L. A. Dawson, and K. K. Brock, "Accumulated dose in liver stereotactic body radiotherapy: Positioning, breathing, and deformation effects," *Int. J. Radiat. Oncol., Biol., Phys.* **83**(4), 1132–1140 (2012).

¹⁴M. Velec, J. L. Moseley, C. L. Eccles, T. Craig, M. B. Sharpe, L. A. Dawson, and K. K. Brock, "Effect of breathing motion on radiotherapy dose accumulation in the abdomen using deformable registration," *Int. J. Radiat. Oncol., Biol., Phys.* **80**(1), 265–272 (2011).

¹⁵H. Delingette, "Toward realistic soft-tissue modeling in medical simulation," *Proc. IEEE* **86**(3), 512–523 (1998).

¹⁶S. L. Delp and J. P. Loan, "A graphics-based software system to develop and analyze models of musculoskeletal structures," *Comput. Biol. Med.* **25**(1), 21–34 (1995).

¹⁷J. Rosen, J. D. Brown, S. De, M. Sinanan, and B. Hannaford, "Biomechanical properties of abdominal organs *in vivo* and postmortem under compression loads," *J. Biomech. Eng.* **130**(2), 021020 (2008).

¹⁸R. M. Satava, "Medical applications of virtual reality," *J. Med. Syst.* **19**(3), 275–280 (1995).

¹⁹R. M. Satava, "Medical virtual reality. The current status of the future medical applications of virtual reality," *Stud. Health Technol. Inf.* **29**(3), 100–106 (1996).

²⁰C. Fiorino *et al.*, "Density variation of parotid glands during IMRT for head-neck cancer: Correlation with treatment and anatomical parameters," *Radiother. Oncol.* **104**(2), 224–229 (2012).

²¹S. Cotin, H. Delingette, and N. Ayache, "Real time volumetric deformable models for surgery simulation," in *Visualization in Biomedical Computing*, edited by K. Höhne and R. Kikinis (Springer Berlin Heidelberg, Germany, 1996), Vol. 1131, pp. 535–540.

²²G. Picinbono, H. Delingette, and N. Ayache, "Non-linear elastic deformable models for real-time surgery simulation," *C. R. Biol.* **325**(4), 335–344 (2002).

²³J. M. Schwartz, M. Denninger, D. Rancourt, C. Moisan, and D. Laurendeau, "Modelling liver tissue properties using a non-linear visco-elastic model for surgery simulation," *Med. Image Anal.* **9**(2), 103–112 (2005).

²⁴E. Basafa and F. Farahmand, "Real-time simulation of the nonlinear visco-elastic deformations of soft tissues," *Int. J. Comput. Assisted Radiol. Surg.* **6**(3), 297–307 (2011).

²⁵R. F. Moyer, W. R. McElroy, J. E. O'Brien, and C. C. Chamberlain, "A surface bolus material for high-energy photon and electron therapy," *Radiology* **146**(2), 531–532 (1983).

²⁶American Society for Testing and Materials (ASTM) D412-98a Rubber Properties in Tension Report, Smithers Scientific Services, Inc., Akron, OH, 2000.

²⁷O. A. Ejofodomi, V. Zderic, and J. M. Zara, "Tissue-mimicking bladder wall phantoms for evaluating acoustic radiation force-optical coherence elastography systems," *Med. Phys.* **37**(4), 1440–1448 (2010).

²⁸R. van Mastrigt, B. L. Coolsaet, and W. A. van Duyl, "Passive properties of the urinary bladder in the collection phase," *Med. Biol. Eng. Comput.* **16**(5), 471–482 (1978).

²⁹F. A. Duck, *Physical Properties of Tissue: A Comprehensive Reference Book* (Academic, New York, NY, 1990).

³⁰S. E. Dahms, H. J. Piechota, R. Dahiya, T. F. Lue, and E. A. Tanagho, "Composition and biomechanical properties of the bladder acellular matrix graft: Comparative analysis in rat, pig and human," *Br. J. Urol.* **82**(3), 411–419 (1998).

³¹<http://www.omega.com/Pressure/pdf/LCKD.pdf>.

³²D. P. Huttenlocher, G. A. Klanderman, and W. A. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 850–863 (1993).

³³C. Pinter, A. Lasso, A. Wang, D. Jaffray, and G. Fichtinger, "SlicerRT: Radiation therapy research toolkit for 3D slicer," *Med. Phys.* **39**(10), 6332–6338 (2012).

³⁴J. Piper, "Evaluation of an intensity-based free-form deformable registration algorithm," *Med. Phys.* **34**, 2353 (2007).

³⁵E. Schreibmann, T. Fox, and I. Crocker, "Dosimetric effects of manual cone-beam CT (CBCT) matching for spinal radiosurgery: Our experience," *J. Appl. Clin. Med. Phys.* **12**(3), 132–141 (2011).

³⁶D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free form deformations," *IEEE Trans. Med. Imaging* **22**(1), 120–128 (2003).

³⁷C. Rubod, M. Brieu, M. Cosson, G. Rivaux, J. C. Clay, L. de Landsheere, and B. Gabriel, "Biomechanical properties of human pelvic organs," *Urology* **79**(4), 968.e17–968.e22 (2012).

³⁸J. L. Sparks and R. B. Dupaix, "Constitutive modeling of rate-dependent stress-strain behavior of human liver in blunt impact loading," *Ann. Biomed. Eng.* **36**(11), 1883–1892 (2008).

³⁹D. Yang, S. R. Chaudhari, S. M. Goddu, D. Pratt, D. Khullar, J. O. Deasy, and I. El Naqa, "Deformable registration of abdominal kilovoltage treatment planning CT and tomotherapy daily megavoltage CT for treatment adaptation," *Med. Phys.* **36**(2), 329–338 (2009).

⁴⁰U. J. Yeo, M. L. Taylor, J. R. Supple, R. L. Smith, L. Dunn, T. Kron, and R. D. Franich, "Is it sensible to 'deform' dose? 3D experimental validation of dose-warping," *Med. Phys.* **39**(8), 5065–5072 (2012).

⁴¹B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.* **17**, 185–204 (1981).

⁴²S. Wognum, S. E. Heethuis, T. Rosario, M. S. Hoogeman, and A. Bel, "Validation of deformable image registration algorithms on CT images of *ex vivo* porcine bladders with fiducial markers," *Med. Phys.* **41**(7), 071916 (11pp.) (2014).

⁴³M. Velec, T. Juang, J. L. Moseley, M. Oldham, and K. K. Brock, "Utility and validation of biomechanical deformable image registration in low-contrast images," *Pract. Radiat. Oncol.* **5**(4), e401–e408 (2015).