Technical note: A physical phantom for assessment of accuracy of deformable alignment algorithms

Rojano Kashani, Martina Hub, Marc L. Kessler, and James M. Balter
Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan 48109

The purpose of this study was to investigate the feasibility of a simple deformable phantom as a QA tool for testing and validation of deformable image registration algorithms. A diagnostic thoracic imaging phantom with a deformable foam insert was used in this study. Small plastic markers were distributed through the foam to create a lattice with a measurable deformation as the ground truth data for all comparisons. The foam was compressed in the superior-inferior direction using a one-dimensional drive stage pushing a flat "diaphragm" to create deformations similar to those from inhale and exhale states. Images were acquired at different compressions of the foam and the location of every marker was manually identified on each image volume to establish a known deformation field with a known accuracy. The markers were removed digitally from corresponding images prior to registration. Different image registration algorithms were tested using this method. Repeat measurement of marker positions showed an accuracy of better than 1 mm in identification of the reference marks. Testing the method on several image registration algorithms showed that the system is capable of evaluating errors quantitatively. This phantom is able to quantitatively assess the accuracy of deformable image registration, using a measure of accuracy that is independent of the signals that drive the deformation parameters. © *2007 American Association of Physicists in Medicine*.
[DOI: 10.1118/1.2739812]

I. INTRODUCTION

Image registration has become an essential part of many stages in the radiation therapy process due to the increase in use of daily imaging for setup adjustment and availability of new scans for replanning during the course of treatment. Therefore, it is important to evaluate and validate the accuracy of image registration algorithms.[1] This need becomes even more pressing with the emerging inclusion of local deformation in the degrees of freedom that image registration algorithms use to account for shape changes due to such factors as breathing, weight gain/loss, and tumor shrinkage over a treatment course.[2–4] Most often, image registration results are evaluated qualitatively using a variety of visual techniques such as split screen displays.[5–7] Previous studies have also evaluated the accuracy of different algorithms for specific body sites, based on anatomical landmarks such as vessel and bronchial bifurcations that were delineated by experts.[8–11] However, precision in identifying these landmarks is greatly dependent on image resolution, slice thickness, and the user. Brock *et al.* reported the average precision in selecting the location of these landmarks to be better than 1 mm, based on repeat measurements done by one individual once a week over a four week period.[8] They also showed a maximum standard deviation of 4.7 mm for one bifurcation in the superior/inferior direction. Using anatomical landmarks for evaluation of image registration results also has the disadvantage that those with enough contrast to be accurately identifiable are likely to contribute significantly to the goodness-of-fit metrics used for alignment and, thus, have an impact on the registration results. Therefore, evaluation of accuracy at these positions could be biased and not necessarily representative of other locations in the volume.

Other studies investigated methods which include simulating known deformations in a given set of images to be compared with what is achieved based on the image registration algorithm. These studies range in complexity from simulation of simple deformations applied to the image through random motion of different points, to more complex biomechanical models that simulate physical tissue deformations using finite-element methods.[12–17] However, simulated images do not have all the noise and other imaging artifacts of real images, which may impact the true alignment accuracy.

In this study we investigate the feasibility of a simple tool for validation of image registration algorithms. We use a deformable phantom, embedded with small but easily identifiable reference marks, to calculate the true deformation as the baseline for comparison of all image registration results. We then remove these markers from the images prior to registration to eliminate their possible impact on alignment. The transformation map from the image registration is then applied to the reference marks and their estimated location is compared to the true location of the markers at the deformed state to quantitatively evaluate accuracy.
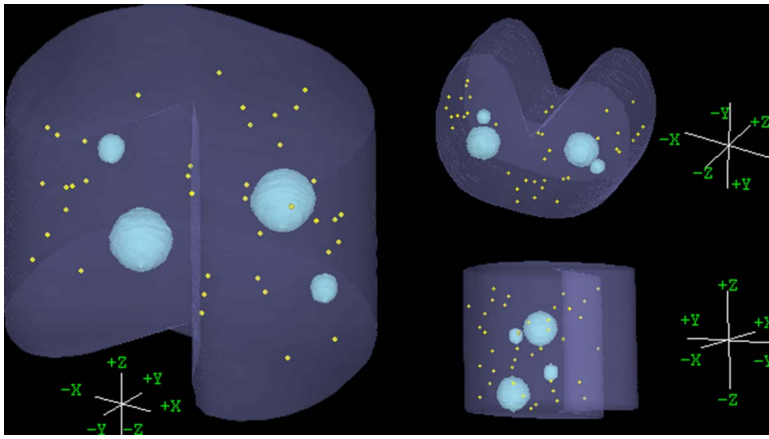
FIG. 1. Volumetric views of the surface of the deformable lung insert, the solid tumor-simulating spheres, and the plastic markers, are shown at 1 cm compression of the foam (inhale). Z: superior/inferior, Y: anterior/posterior, X: right/left directions.

II. METHODS AND MATERIALS

A. Data collection

We previously described a deformable phantom for use in imaging studies which consists of a diagnostic thoracic phantom as the outer shell, with a deformable foam insert that was compressed using a one-dimensional drive stage to simulate different breathing states.[18,19] In this study, the foam insert was modified by placing 48 small (2.5 mm diam) plastic markers at different locations in the foam (Fig. 1). The phantom was then imaged at 1 cm (inhale) and 4 cm (exhale) compressions of the foam insert, using a commercial CT scanner (HighSpeed, General Electric, Milwaukee, WI). Axial images were acquired with 1 mm slice thickness and separation.

B. Data processing

For each image set, the position of the center of each marker was manually measured by locating the central slice (longitudinal position) and then finding the center of the marker on that slice (axial position). This introduces an uncertainty in the measurement of the longitudinal position that is limited by the slice thickness of 1 mm.

In order to evaluate the accuracy of the manual measurement of marker location in the axial plane, a random set of ten markers were chosen and the measurement of their position was repeated three times by a single observer. The standard deviation in the measurements for each marker was cal-

culated and averaged over all markers as a measure of uncertainty in our estimate of the true motion and deformation.

Once the position of the markers was identified on both image sets, the markers were digitally removed from the images prior to image registration. The voxel values in the marker locations were replaced by intensity values of neighboring voxels outside of the markers. Next, Gaussian smoothing with a kernel width of 20 voxels was applied to the intensities of the voxels in the marker locations. This would eliminate the possible impact of reference marks on the registration outcome. Figure 2 shows an original axial image on the left and the image without the markers on the right.

C. Image registration

Image registration was performed on the modified images without the markers, using different in-house algorithms such as rigid and affine transforms as well as thin-plate splines and B-splines.[9] One set of images (1 cm compression or inhale) was chosen as the reference in all cases and for each algorithm several combinations of the relevant parameters (resolution, knot spacing, the number of control points) were tested to get visually acceptable results. Once the registration was done, the resulting transform was applied to the position of the reference marks in the reference image set to estimate their location in the other image set. This estimate
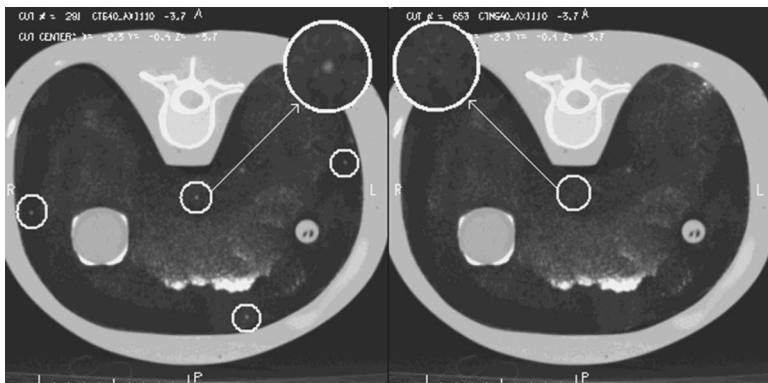

FIG. 2. Axial view of a plane with four markers before (left) and after (right) removal of the markers from the image.

TABLE I. Error in estimation of marker position based on example alignment results from different image registration algorithms.

		RL[a] (cm)	AP[b] (cm)	SI[c] (cm)	3-D distance (cm)
Affine	Average	−0.01	0.00	0.05	0.38
	Stdev[d]	0.04	0.04	0.44	0.22
	Max[e]	−0.12	−0.13	0.90	0.90
B-splines	Average	−0.02	−0.01	0.05	0.18
	Stdev[d]	0.08	0.06	0.22	0.16
	Max[e]	−0.42	0.19	0.67	0.81
Thin-plate splines	Average	−0.07	−0.15	−0.14	0.37
	Stdev[d]	0.12	0.19	0.28	0.19
	Max[e]	−0.56	−0.58	−0.74	0.75

[a]RL: right/left.
[b]AP: anterior/posterior.
[c]SI: superior/inferior.
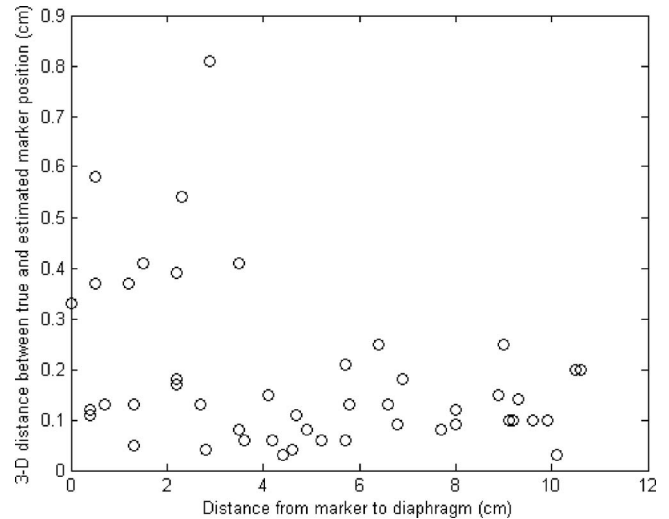[d]Stdev: standard deviation.
[e]Max: maximum.

FIG. 3. The 3D distance between the real marker position and the estimated position based on B-spline transformation, shown as a function of the longitudinal position of the marker relative to the diaphragm. It can be seen that the error in the registration decreases in regions farther away from the diaphragm with less deformation.

was then compared to the actual position of the markers and the accuracy of the image registration was evaluated.

To determine if, in the case of the alignment methods tested, a bias would be introduced by the markers in the image, a second set of alignments was performed on the unedited image volumes.

III. RESULTS

A. Reference point selection accuracy

The accuracy of manual reference mark identification, which was determined by repeat measurements of the location of ten randomly chosen reference marks by a single observer, was found to be better than 0.2 mm (σ) in all directions. The maximum standard deviation of any single measured point location was 0.3, 0.4, and 0.6 mm in right/left (RL), anterior/posterior (AP), and superior/inferior (SI) directions. It should be noted that this error applies to identification of reference marks on each of the two image sets. Therefore, the overall accuracy in the measurement of the true motion of the reference marks is 0.4, 0.5, and 0.8 mm in RL, AP, and SI directions.

B. Image registration

The results for the average and standard deviation as well as maximum differences between the manually measured locations of the reference markers and those locations predicted by different alignment algorithms are shown in Table I. The average 3D distance between the markers' estimated and actual locations was also calculated (Fig. 3). Comparison of image registration on datasets with and without the markers (with the markers removed) showed no significant difference in the results of the image registration accuracy.

IV. DISCUSSION

In this study we demonstrated a simple QA tool for validation of image registration algorithms. This method, unlike most others currently used, does not rely on sample patient data, selection of anatomical landmarks, or simulation of deformations on image sets. Instead it applies a known deformation to a simple geometry and quantitatively evaluates the outcome of the registration by comparing the measured and estimated location of a series of points with a known accuracy. Using reference marks distributed in the phantom, the accuracy can be calculated in all locations, which is not possible with patient data with a limited number of landmarks that can be identified. The ability to remove the reference marks prior to registration can eliminate any bias in the measurement of accuracy at these points that can drive the registration and, therefore, are likely to be more accurately registered. Although in this study the size of the reference marks that we used was small enough that no significant difference was observed, it is still possible that variations in the amount of local signal and deformation may lead to bias and, thus, it is advised that image editing for marker removal be maintained for future studies.

This method can complement the qualitative assessment of image registration results achieved through visualization techniques. It can quantify the errors that are observed in the matched images as well as identify errors in regions where no mismatch was detected by the observer, or to validate the observer's perception of the magnitude of mismatch.

It should be noted that this study does not attempt to compare different image registration algorithms, and the data presented here for accuracy should not be interpreted as a measure of superiority of one algorithm over the other. In order to compare different image registration algorithms, a comprehensive study is necessary to characterize each algo-

rithm individually first. Although this is a possible area of investigation using this phantom, it was not done for this study.

ACKNOWLEDGMENT

[1] M. L. Kessler, "Image registration and data fusion in radiation therapy," Br. J. Radiol. **79**, S99–S108 (2006).

[2] W. Y. Song, E. Wong, G. S. Bauman, J. J. Battista, and J. Van Dyke, "Dosimetric evaluation of daily rigid and nonrigid geometric correction strategies during on-line image-guided radiation therapy (IGRT) of prostate cancer," Med. Phys. **34**, 352–365 (2007).

[3] E. K. Hansen, M. K. Bucci, J. M. Quivey, V. Weinberg, and P. Xia, "Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer," Int. J. Radiat. Oncol., Biol., Phys. **64**, 355–362 (2006).

[4] R. Mohan, X. Zhang, H. Wang, Y. Kang, X. Wang, H. Liu, K. K. Ang, D. Kuban, and L. Dong, "Use of deformed intensity distributions for on-line modification of image-guided IMRT to account for interfractional anatomic changes," Int. J. Radiat. Oncol., Biol., Phys. **61**, 1258–1266 (2005).

[5] U. Malsch, C. Thieke, P. E. Huber, and R. Bendl, "An enhanced block matching algorithm for fast elastic registration in adaptive radiotherapy," Phys. Med. Biol. **51**, 4789–4806 (2006).

[6] J. M. Fitzpatrick, D. L. G. Hill, Y. Shyr, J. West, C. Studholme, and C. R. Maurer, "Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain," IEEE Trans. Med. Imaging **17**, 571–585 (1998).

[7] J. R. McClelland, J. M. Blackall, S. Tarte, A. C. Chandler, S. Hughes, S. Ahmad, D. Landau, and D. Hawkes, "A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy," Med. Phys. **33**, 3348–3358 (2006).

[8] K. K. Brock, M. B. Sharp, 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).

[9] M. M. Coselmon, J. M. Balter, D. L. McShan, and M. L. Kessler, "Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines," Med. Phys. **31**, 2942–2948 (2004).

[10] E. Rietzel and G. T. Y. Chen, "Deformable registration of 4D computed tomography data," Med. Phys. **33**, 4423–4430 (2006).

[11] J. P. Voroney, K. K. Brock, C. Eccles, M. Haider, and L. A. Dawson, "Prospective comparison of computed tomography and magnetic resonance imaging for liver cancer delineation using deformable image registration," Int. J. Radiat. Oncol., Biol., Phys. **66**, 780–791 (2006).

[12] J. A. Schnabel, C. Tanner, A. D. Castellano-Smith, A. Degenhard, M. O. Leach, D. R. Hose, and D. L. G. Hill, "Validation of nonrigid image registration using finite-element methods: Application to breast MR images." IEEE Trans. Med. Imaging **22**, 238–247 (2003).

[13] Y. Y. Chou and O. Skrinjar, "Ground truth data for validation of nonrigid image registration algorithms," IEEE Int. Symp. Biomed. Imag. Macro to Nano **1**, 716–719 (2004).

[14] T. Guerrero, G. Zhang, T. Huang, and K. Lin, "Intrathoracic tumor motion estimation from CT imaging using the 3D optical flow method," Phys. Med. Biol. **49**, 4147–4161 (2004).

[15] H. 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).

[16] H. Wang, L. Dong, M. F. Lii, R. de Crevoisier, R. Mohan, J. D. Cox, D. A. Kuban, and R. Cheung, "Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy," Int. J. Radiat. Oncol., Biol., Phys. **61**, 725–735 (2005).

[17] V. Walimbe and R. Shekhar, "Automatic elastic image registration by interpolation of 3D rotations and translations from discrete rigid-body transformations," Med. Image Anal **10**, 899–914 (2006).

[18] R. Kashani, K. Lam, D. W. Litzenberg, and J. M. Balter, "Technical note: A deformable phantom for dynamic modeling in radiation therapy," Med. Phys. **34**, 199–201 (2007).

[19] D. W. Litzenberg, S. W. Hadley, N. Tyagi, J. M. Balter, R. K. Ten Haken, and I. J. Chetty, "Synchronized dynamic dose reconstruction," Med. Phys. **34**, 91–102 (2007).