Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2013 June 1; 86(2): 372–379. doi:10.1016/j.ijrobp.2012.12.023.

Evaluation of Four-Dimensional Computed Tomography to Four-Dimensional Cone-Beam Computed Tomography Deformable Image Registration for Lung Cancer Adaptive Radiation Therapy

Salim Balik, Ph.D.^{*}, Elisabeth Weiss, M.D.^{*}, Nuzhat Jan, M.D.^{*}, Nicholas Roman, M.D.^{*}, William C. Sleeman, M.S.^{*}, Mirek Fatyga, Ph.D.^{*}, Gary E. Christensen, D.Sc.[†], Cheng Zhang, M.S.[‡], Martin J. Murphy, Ph.D.^{*}, Jun Lu, Ph.D.[§], Paul Keall, Ph.D.^{||}, Jeffrey F. Williamson, Ph.D.^{*}, and Geoffrey D. Hugo, Ph.D.^{*}

^{*}Department of Radiation Oncology, Virginia Commonwealth University, Richmond, Virginia

[†]Department of Electrical and Computer Engineering and Radiation Oncology, The University of Iowa, Iowa City, Iowa

[‡]Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, Iowa

[§]Department of Radiation Oncology, SUNY Upstate Medical University, Syracuse, New York

^{||}Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia

Abstract

Purpose—To evaluate two deformable image registration (DIR) algorithms for the purpose of contour mapping to support image guided adaptive radiotherapy with four-dimensional cone beam computed tomography (4DCBCT).

Methods and Materials—One planning 4D fan-beam CT (4DFBCT) and 7 weekly 4DCBCT scans were acquired for 10 locally advanced non-small cell lung cancer patients. The gross tumor volume (GTV) was delineated by a physician in all 4D images. End-of-inspiration phase planning 4DFBCT was registered to the corresponding phase in weekly 4DCBCT images for day-to-day registrations. For phase-to-phase registration, the end-of-inspiration phase from each 4D image was registered to the end-of-expiration phase. Two DIR algorithms—small deformation inverse consistent linear elastic (SICLE) and Insight Toolkit diffeomorphic demons (DEMONS)—were evaluated. Physician-delineated contours were compared to the warped contours by using the Dice similarity coefficient (DSC), average symmetric distance (ASD), false positive and false negative indices. The DIR results are compared to rigid registration of tumor.

Results—For day-to-day registrations, the mean DSC was 0.75 ± 0.09 with SICLE, 0.70 ± 0.12 with DEMONS, 0.66 ± 0.12 with rigid-tumor registration and 0.60 ± 0.14 with rigid-bone registration. Results were comparable to intra-observer variability calculated from phase-to-phase registrations as well as measured inter-observer variation for one patient. SICLE and DEMONS, when compared to rigid-bone (4.1 mm) and rigid-tumor (3.6 mm) registration, respectively

© 2013 Elsevier Inc. All rights reserved.

Reprint requests to: Geoffrey D. Hugo, Ph.D., Department of Radiation Oncology, Virginia Commonwealth University, 401 College Street, PO Box 980058, Richmond, VA 23298. Tel: (804) 628-3457; gdhugo@vcu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflicts of Interest Notification

Dr. Christensen hold a patent related to the SICLE registration algorithm.

reduced the ASD to 2.6 and 3.3 mm. On average, SICLE and DEMONS increased the DSC to 0.80 and 0.79 respectively, compared to rigid-tumor (0.78) registrations for 4DCBCT phase-to-phase registrations.

Conclusions—DIR achieved comparable accuracy to reported inter-observer delineation variability and higher accuracy than rigid-tumor registration. DIR performance varied with the algorithm and the patient.

Introduction

Radiation therapy planning is conventionally based on the initial planning computed tomography (CT) image, but treatment of non-small cell lung cancer (NSCLC) is complicated by the fact that breathing, tumor regression or growth, and other pathological changes lead to intrafractional and interfractional variations in a patient's anatomy. With the advent of cone-beam CT (CBCT) - guided radiotherapy, it has become possible to image patients before and during the treatment, potentially reducing uncertainties related to setup errors. CBCT has made continuous monitoring of the patient's anatomical changes during treatment more readily available, thus enabling image-guided adaptive radiotherapy (IGART). It has been shown that adaptation of radiotherapy to anatomical changes allows dose to be escalated to the gross tumor volume (GTV), potentially improving local control (1, 2).

A key requirement for efficient IGART is accurate deformable image registration (DIR) that tracks voxel-to-voxel from the initial clinical planning image to the images acquired over the course of treatment. In free-breathing NSCLC patients, this is complicated by the fact that both CBCT and FBCT images are 4D, consisting of as many as 10 3D phase-specific image sets. DIR allows the transfer of important clinical data (such as contours of targets and organs at risk) from one CT image to another, and enables the delivered dose to be mapped to a consistent reference frame for evaluation of cumulative dose. DIR involving 4DCBCT images may be difficult due to view-aliasing artifacts, which lead to reduced contrast and CT number accuracy and structural artifact such as streaking.

Partly due to these perceived challenges, studies of DIR of 4DCBCT have been limited (2–10). Previous work has focused on DIR between respiratory phases of a single 4-dimensional (4D) CT image. For IGART, however, DIR between images acquired on different days is required. This study evaluates the use of two DIR algorithms to support IGART with weekly 4DCBCT. DIR accuracy is measured in planning fan-beam CT (FBCT) to weekly CBCT registrations, and registrations between respiratory phases of 4DFBCT and 4DCBCT images. DIR results are compared to rigid registration results based on bony anatomy and region of interest (ROI) around the tumor.

Methods and Materials

Patient data

In accordance with a protocol approved by the local institutional review board, 1 planning 4DFBCT and 7 weekly 4DCBCT images were acquired from 10 patients with locally advanced NSCLC. The 4D images, which were binned by phase, consisted of a 3-dimensional (3D) image set for each of the 10 respiratory phases. During image acquisition, audiovisual biofeedback was used to coach the patient and reduce irregularities in respiration.

Planning 4D-FBCT images were acquired using a 16-slice, helical CT scanner (Brilliance Big Bore, Philips Medical Systems) with a slice thickness of 3 mm and 512×512 axial

resolution (~1 mm pixel size). The 4D-CBCT images were acquired on a commercial CBCT system (On- Board Imager™, Varian Medical Systems, Inc.). Using in-house hardware and software modifications, an external surrogate (Real-time Position Management, Varian Medical Systems, Inc.) was integrated into the CBCT acquisition system to stamp each CBCT projection with the surrogate respiratory signal. Approximately 2500 projections were acquired over a period of 8- 10 minutes in half-fan mode with half bow-tie filter and with x-ray tube settings of 125 kVp, 20 mA, and 20 ms in a single 360° slow gantry arc. The 512×512×50 CBCT images (0.88 mm and 3 mm pixel size in the inplane and longitudinal axes respectively) were sorted into phases and reconstructed with in-house FDK reconstruction software (11).

GTV was contoured by a physician on end-of-inhale and end-of-exhale phases of the 4DFBCT and 4DCBCT for 8 patients. To assess inter-observer contouring variation, for one patient (P8), GTV in the end-of-inhale phase images were contoured by two physicians. GTV contours were then converted into a 3D triangular mesh with the maximum number of vertices ranging from 5000 to 42,000 using the research version of Pinnacle (version 8.1, Philips Medical Systems).

Registration procedure

Two DIR algorithms were evaluated in day-to-day and phase-to-phase registrations: the small deformation, inverse consistent linear elastic (SICLE) algorithm (12), and the Insight Toolkit v3.18 implementation of the diffeomorphic demons (DEMONS) algorithm (13). SICLE is regularized using a linear elastic constraint and is inverse consistent, while DEMONS is widely available, less regularized, and is a faster algorithm. The details of the deformable registration algorithms are given in Appendix I. The DIR results were compared to the rigid registration of tumor and the inter-observer variation (for one patient).

Phase-to-phase registrations were performed for 8 patients. The end-of-inhale phase was registered to end-of-exhale phase for both 4DFBCT and 4DCBCT images. In day-to-day registration, the end-of-inhale phase of the 4DFBCT image was registered to the corresponding phase of each 4DCBCT image. Day-to-day registrations were performed in two ways: FBtoCB, with the FBCT image (moving) registered to the CBCT image (fixed); and CBtoFB, with the CBCT image (moving) registered to the FBCT image (fixed).

All images were rigidly registered to bony anatomy using in-house software (3 translations and 3 rotations) before DIR (RIGID-BONE). The field of view (FOV) of the FBCT images in the superior-inferior direction are much larger than the CBCT images which pose a problem in CBtoFB registrations since part of the anatomy that exists in FBCT image is not visible in the CBCT image. The superior-inferior FOV of the FBCT was cropped to match that of the CBCT image for CBtoFB DEMONS registrations (The FBCT FOV was not changed for DEMONS FBtoCB registrations). SICLE algorithm requires two identical size images so the FOV was adjusted similarly. Finally, the images were resized to 256×256 in the axial plane.

In rigid-tumor registration, the images were rigidly registered only using the region of interest defined by minimum bounding box around the tumor with 1 cm margin (RIGID-TUMOR).

Evaluation

Registration evaluation is a difficult task since there is no “ground truth” information that maps the points from one image to another. Physician-delineated structures were used as the reference standard. The displacement vector field (DVF) resulting from the registration maps coordinates in the fixed image domain into the moving image domain. The physician-

delineated structures were mapped to the other image using the estimated DVF and compared to the contours delineated on the image (Fig. 1). DIR accuracy was quantified by both volumetric and surface congruence measures.

Volumetric

Three volumetric evaluation measures were used in this work. The Dice similarity coefficient (DSC)(14) measures the overlap of two segmentations A and B and is defined as

$$DSC = \frac{2|A \cap B|}{|A| + |B|}. \quad (1)$$

A value of 1 indicates perfect overlap and 0 indicates no overlap. The false positive (FP) index measures the fraction of deformed structure not overlapping with the manual structure,

$$FP = \frac{|A| - |A \cap B|}{|A|}, \quad (2)$$

where A is the deformed structure and B is the manual structure. A value of 0 means that the deformed structure is completely confined within the manual structure and $|A| \leq |B|$. The false negative (FN) index measures the fraction of the manual structure that is not overlapping with the warped structure:

$$FN = \frac{|B| - |A \cap B|}{|B|}. \quad (3)$$

A value of 0 means that the deformed structure confines the manual structure completely and $|B| \leq |A|$. Smaller FP and FN values indicate a better match of the deformed and manual structures.

Surface congruence

Pinnacle was used to generate a triangular mesh of vertices for each segmentation. For every vertex point in the deformed mesh (A), the closest vertex point in the manual mesh (B) was found and the Euclidean distance was calculated. The same procedure was repeated from B to A. The average symmetric distance (ASD) (15) and 90% confidence interval ($D_{90\%}$) were used to evaluate the surface congruence.

Statistical methods

Statistical significance was measured using Student paired *t*-test if the Shapiro-Wilk normality test passed ($P > .050$), otherwise the Wilcoxon Signed Rank test was used. The results were considered significant for $P < .05$.

Results

Progressive GTV shrinkage was observed during treatment for 9 of the 10 patients, with an average GTV reduction of 40% from first to last CBCT (Fig 2).

Day-to-day DSC results are shown in Fig. 3a. The average improvement in DSC with DIR compared to RIGID-TUMOR registration were $14.9 \pm 18.7\%$ in SICLE FBtoCB, $18.2 \pm 21.9\%$ in SICLE CBtoFB, $4.4 \pm 21.0\%$ in DEMONS FBtoCB, and $15.3 \pm 27.1\%$ in DEMONS CBtoFB registrations (Table 1).

In general, both DIR algorithms showed a significant decrease in FP and FN indices compared to RIGID-TUMOR as shown in Table 1.

The results of the surface congruence analysis (ASD and $D_{90\%}$) are consistent with the mean DSC results (Table 1). The patient-specific surface congruence results show large differences compared to the corresponding DSC results (Fig. 3). In addition, the ASD results show a correlation with planning FBCT GTV volume (Pearson correlation coefficient [PCC] = 0.88, $P < .001$ for SICLE CBtoFB) and there is no such correlation for DSC (PCC = 0.14, $P = .7$). The correlation can be seen in Fig. 3b as the patients are ordered by planning FBCT GTV volume. The spread of the minimum distances are shown in Fig. 4.

Phase-to-phase registration results are shown in (Table 3). SICLE improved the mean DSC results for both the CBCT and FBCT images, whereas DEMONS produced similar results compared to RIGID-TUMOR registration.

For patient P8, a mean DSC of 0.82 ± 0.03 is calculated using two physician contours which is comparable to the DIR results (SICLE CBtoFB 0.82 ± 0.03 , DEMONS CBtoFB 0.81 ± 0.04) for this patient.

Discussion

The mean DSC from deformable phase-to-phase registrations of 4DFBCT images generally agrees with previous contour propagation studies for 4D lung datasets (3–8). The phase-to-phase results show smaller mean DSC values for the CBCT images than the FBCT images which can be attributed to manual contouring variations (16). This results in inconsistent contours in FBCT and CBCT, adversely affecting day-to-day evaluations. In the pelvis, Weiss *et al.* (17) showed that inter-observer variations in detecting the soft tissue boundary is larger on CBCT than it is on FBCT.

Day-to-day DIR registration results (Table 1) show generally significant improvements compared to RIGID-TUMOR registration. The agreement between the physician and DIR generated contours (with a mean DSC value of 0.76 ± 0.08) is classified as “substantial” category of DSC (0.61–0.80 in kappa statistics) (18). The results are also similar to reported DSC of 0.76 with the symmetric forces Demons algorithm for weekly FBCT to CBCT registration in head and neck patients (19).

Achieving contour propagation error similar to physician inter-observer variability is a goal for DIR-assisted contour propagation. One limitation of this study is not having inter-observer variability measured in all patients. On the other hand, the phase-to-phase CBCT RIGID-TUMOR results reflect the underlying intra-observer contouring variations since no tumor shrinkage is involved. The effect of GTV deformations like compression or stretching during breathing is estimated to be small compared to intra-observer variations since measured GTV volumes at end-of-inhale were not consistently bigger than at end-of-exhale in this study. SICLE CBtoFB registration DSC results are similar to phase-to-phase CBCT RIGID-TUMOR results (Mann-Whitney rank test, $p = 0.273$). Thus, autogenerated contours appear to be reasonable for IGART but may need additional physician review before clinical use, or may be used to aid in the contouring process.

Existing inter-observer or intra-observer variability studies generally use different comparison measures than were used in this study, and they do not use 4DCBCT, which makes direct comparison to our results difficult. In a recent CBCT-based delineation study of stereotactic lung GTV by Altorjai *et al.*, (16), a mean inter-observer concordance index (CI) of 0.64 ± 0.07 was reported, which is similar to the CI value for SICLE CBtoFB registration (0.62 ± 0.11) that we observed.

The surface similarity results showed improvement in comparison to RIGID-TUMOR registration. The ASD and $D_{90\%}$ results are comparable to the DIR and inter-observer results reported by Pevsner *et al.* (mean distance and 90% confidence interval: 2.6 and 5.1 mm respectively for DIR, and 2.1 and 3.7 mm for inter-observer) (4). In a DIR evaluation study of NSCLC, Guckenberger *et al.* (9) reported mean surface distance of 2.9 ± 1.2 using 6 points around the GTV. More recently, Hugo *et al.* (10) reported a DIR accuracy of 3.1 mm when measured near the CTV borders using vascular and bronchial bifurcation landmarks.

The differences in the surface congruence and volume overlap measures shown in Fig. 3 can be explained by differences in GTV volumes and locations among the patients P1, P2 and P3. For smaller GTVs, a given level of surface discrepancy results in large DSC changes. For example, for two spherical concentric volumes with surface differences of 2 mm, 10 cm³ and 150 cm³ volumes have DSCs of 0.79 and 0.91, respectively. As illustrated by P9 and P10, a DSC of 0.79 results in a mean coincidence error of 5 mm, due to the large size of these tumors. Because of this inverse relation between the volume and surface overlap measure, the following suggestions are made for future contour evaluation studies: (1) use DSC for small structures, (2) ASD or similar measures for larger structures, or (3) a combined measure that uses volumedependent relative weights to DSC and ASD. Furthermore, for patients P1, P2, and P3, the GTV locations in high contrast regions with clear tumor boundaries might cause more accurate deformation of GTV and consistent physician delineation (17).

Since DSC and ASD rely on physician-delineated contours as the “ground truth” for DIR validation, inter-observer and intra-observer contouring variability should be considered (5, 6, 16, 17). Any delineation inconsistencies between the FBCT and CBCT images adversely affect estimated DIR accuracy. For example, the relatively large increases in GTV volume (~40%; see Fig 2) for patients P3 and P4 between the planning FBCT and initial CBCT are more likely due to delineation inconsistencies.

Examples of variations in DIR-generated GTV contours are shown from 3 patients in Fig. 1. The figure shows that the deformed contour matched well in general for patient 2, where the lesion is surrounded by lung tissue and has a clear border. For patient P9, the medial aspect of the tumor is difficult to distinguish from normal mediastinal structures, potentially resulting in delineation inconsistencies. As illustrated by Fig. 1; bottom row (P10), DEMONS deformably mapped contours often have irregular shapes with undulating boundaries, indicating the need to smooth the underlying PDFs prior to clinical use.

Although CBtoFB registrations generally produced significantly better results than FBtoCB for SICLE (paired test $P < .001$; unpaired test $P = 0.159$), the clinical effect of the difference would likely be small. This result is an artifact of the evaluation methods, rather than the image order during registration, since SICLE is a symmetric DIR algorithm and imposes consistency upon the forward and inverse transforms. Mapping the larger tumor volume from the planning FBCT image into the daily CBCT image with the FBtoCB registration results in a smaller volume due to tumor shrinkage. The smaller tumor volume for both the mapped and delineated structures results in a lower DSC. Therefore, CBtoFB registrations appear more accurate than FBtoCB. For DEMONS, the discrepancy between measures for CBtoFB and FBtoCB was larger than for SICLE, and so may not be solely explained by tumor shrinkage. Because inverse consistency was not imposed on these registrations, potentially the order of registration may be more important for this DEMONS implementation.

Conclusion

The accuracy of the SICLE and DEMONS algorithms was evaluated for free-breathing NSCLC patients using physician delineated contours as reference information. Although the DIR algorithms showed variability with physician-contours, they consistently outperformed rigid registration of tumor and yielded results similar to intra-observer variations. On average, SICLE performed somewhat better than DEMONS and exhibited superior consistency between forward and inverse contour mapping.

Acknowledgments

We would like to thank Josh Evans, Douglas Vile, John C. Ford, Scott Robertson, Dandan Zheng, Baoshe Zhang, Emily Brackbill and Chris Barteel for assisting in 4DFBCT and 4DCBCT data acquisition, and Kevin T. Hogan for editorial assistance with the manuscript. This work was supported by grant number P01CA116602 by the National Cancer Institute.

References

1. Fox J, Ford E, Redmond K, et al. Quantification of Tumor Volume Changes During Radiotherapy for Non-Small-Cell Lung Cancer. *International Journal of Radiation Oncology* Biology* Physics*. 2009; 74:341–348.
2. Guckenberger M, Wilbert J, Richter A, et al. Potential of adaptive radiotherapy to escalate the radiation dose in combined radiochemotherapy for locally advanced nonsmall cell lung cancer. *International Journal of Radiation Oncology* Biology* Physics*. 2011; 79:901–908.
3. Ezhil M, Choi B, Starkschall G, et al. Comparison of rigid and adaptive methods of propagating gross tumor volume through respiratory phases of four-dimensional computed tomography image data set. *International Journal of Radiation Oncology* Biology* Physics*. 2008; 71:290–296.
4. Pevsner A, Davis B, Joshi S, et al. Evaluation of an automated deformable image matching method for quantifying lung motion in respiration-correlated CT images. *Medical Physics*. 2006; 33:369. [PubMed: 16532942]
5. Speight R, Sykes J, Lindsay R, et al. The evaluation of a deformable image registration segmentation technique for semi-automating internal target volume (ITV) production from 4DCT images of lung stereotactic body radiotherapy (SBRT) patients. *Radiotherapy and Oncology*. 2011; 98:277–283. [PubMed: 21257217]
6. van Dam IE, van Sörnsen de Koste JR, Hanna GG, et al. Improving target delineation on 4-dimensional CT scans in stage I NSCLC using a deformable registration tool. *Radiotherapy and Oncology*. 2010; 96:67–72. [PubMed: 20570381]
7. Wijesooriya K, Weiss E, Dill V, et al. Quantifying the accuracy of automated structure segmentation in 4D CT images using a deformable image registration algorithm. *Medical Physics*. 2008; 35:1251. [PubMed: 18491517]
8. Weiss E, Wijesooriya K, Ramakrishnan V, et al. Comparison of Intensity-Modulated Radiotherapy Planning Based on Manual and Automatically Generated Contours Using Deformable Image Registration in Four-Dimensional Computed Tomography of Lung Cancer Patients. *International Journal of Radiation Oncology* Biology* Physics*. 2008; 70:572–581. e572.
9. Guckenberger M, Baier K, Richter A, et al. Evolution of surface-based deformable image registration for adaptive radiotherapy of non-small cell lung cancer (NSCLC). *BioMed Central*. 2009
10. Hugo GD, Weiss E, Badawi A, et al. Localization Accuracy of the Clinical Target Volume During Image-Guided Radiotherapy of Lung Cancer. *International Journal of Radiation Oncology* Biology* Physics*. 2011; 81:560–567.
11. Lu J, Guerrero TM, Munro P, et al. Four-dimensional cone beam CT with adaptive gantry rotation and adaptive data sampling. *Medical Physics*. 2007; 34:3520–3529. [PubMed: 17926955]
12. Christensen GE, Song JH, Lu W, et al. Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. *Medical Physics*. 2007; 34:2155–2163. [PubMed: 17654918]

13. Vercauteren, T.; Pennec, X.; Perchant, A., et al. Non-parametric Diffeomorphic Image Registration with the Demons Algorithm Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007. In: Ayache, N.; Ourselin, S.; Maeder, A., editors. Vol. Vol 4792. Berlin / Heidelberg: Springer; 2007. p. 319-326.
14. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945; 26:297–302.
15. Heimann T, Van Ginneken B, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *Medical Imaging, IEEE Transactions on*. 2009; 28:1251–1265.
16. Altorjai G, Fotina I, Lütgendorf-Caucig C, et al. Cone-Beam CT-Based Delineation of Stereotactic Lung Targets: The Influence of Image Modality and Target Size on Interobserver Variability. *International journal of radiation oncology, biology, physics*. 2011
17. Weiss E, Wu J, Sleeman W, et al. Clinical Evaluation of Soft Tissue Organ Boundary Visualization on Cone-Beam Computed Tomographic Imaging. *International Journal of Radiation Oncology*Biography*Physics*. 2010; 78:929–936.
18. Allozi R, Li XA, White J, et al. Tools for consensus analysis of experts' contours for radiotherapy structure definitions. *Radiotherapy and Oncology*. 2010; 97:572–578. [PubMed: 20708285]
19. Hou J, Guerrero M, Chen W, et al. Deformable planning CT to cone-beam CT image registration in head-and-neck cancer. *Medical Physics*. 2011; 38:2088. [PubMed: 21626941]

Appendix I

DEMONS algorithm

The DEMONS algorithm assumes that homologous contours in the fixed and moving images are represented with the same intensity. The contours in the fixed image are considered semipermeable membranes where the effectors, called “demons,” are scattered along the contours and can identify points in the moving contours as “inside” or “outside” points. The orientation and magnitude of the displacement of the points are calculated with optical flow equations. The registration iteratively minimizes the sum of squared intensity differences between the images. To avoid unrealistic results, a Gaussian filter with a standard deviation of 1 mm was applied at each iteration.

Since CBCT and FBCT images have different intensities, histogram matching with 1024 histogram levels and 7 histogram matching points were used. All registrations were performed with 200 iterations (which achieved convergence in all cases) at 256×256 in-plane resolution.

SICLE algorithm

For SICLE, the forward (h) and reverse (g) transformations are jointly estimated such that an inverse consistency error is minimized when $h = g^{-1}$. This algorithm iteratively minimizes the following symmetric cost function

$$C_{SIM}(T \circ h, S) + C_{SIM}(S \circ g, T) + C_{INV}(h, g) + C_{REG}(h) + C_{REG}(g). \quad (4)$$

where $T \circ h$ is image T deformed with the forward transformation h , and $S \circ g$ is image S deformed with the reverse transformation g . C_{SIM} is a squared intensity difference cost, C_{INV} is the inverse consistency cost, and C_{REG} is an elastic regularization term. The SICLE algorithm produces diffeomorphic transformations by ensuring that the Jacobian of the transformation does not go negative. A multiresolution registration procedure was used to reduce local registration errors and to increase the convergence rate. The algorithm was run at $1/8^{\text{th}}$ (32×32), $1/4^{\text{th}}$ (64×64), $1/2$ (128×128), and full (256×256) resolution for 500, 500, 100 and 20 iterations, respectively.

Summary

The accuracy in mapping of gross tumor volumes from four-dimensional planning CT to four-dimensional cone beam CT images using two deformable image registration algorithms was evaluated. Although accuracy relative to physician delineation was variable across patients for both algorithms, both consistently outperformed rigid registration and yielded accuracy consistent with inter-observer contouring variation. These algorithms are strong candidates to use in future planned clinical adaptive radiation therapy trials.

Fig. 1.

Three examples of variations in day-to-day deformable image registration (DIR) generated gross tumor volume (GTV) contours. Images from patients 2 and 9 were registered with SICLE, and images from Patient 10 were registered with DEMONS. (Top Row) Superimposed planning fan-beam CT (FBCT) and last fraction cone beam CT (CBCT) images and associated manual contours aligned via rigid registration. (Middle Row) Overlay of deformed FBCT contour (blue) and the last fraction CBCT contour (green) on CBCT image. (Bottom Row) Deformed weekly CBCT contours superimposed on the planning FBCT image (result of FBtoCB registrations). The physician delineated FBCT GTV contour is represented by the blue color-wash.

Fig. 2.
GTV volumes measured from physician contours; (a) absolute, (b) relative.

Fig. 3. Dice similarity coefficient (DSC) and average symmetric distance (ASD) results for day-to-day registrations with SICLE and DEMONS DIR algorithms compared to rigid registration. Patients are ordered by increasing GTV as measured on 4DFBCT. (a) DSC results. (b) ASD results.

Fig. 4.

Day-to-day registration whiskers box plots of the minimum distances between manual mesh and mesh generated with (a) the rigid-bone registration, (b) the rigid-tumor registration (c) the SICLE CBtoFB, and (d) the DEMONS CBtoFB algorithms. The boundary of the box indicates the 25th and 75th percentiles; the black line within the box (lower) marks the median and the red line marks the mean; error bars indicate the 90th and 10th percentiles; circles indicate the 5th and 95th percentiles.

Table 1

Mean values of the evaluation measures for day-to-day registrations

Algorithm	DSC ± SD	FP ± SD	FN ± SD	ASD ± SD (mm)	D _{90%} ± SD (mm)
RIGID-BONE	0.60 ± 0.14	0.45 ± 0.16	0.30 ± 0.15	4.1 ± 1.5	8.5 ± 3.5
RIGID-TUMOR	0.66 ± 0.12 <i>P_{RB} < .001 *</i>	0.40 ± 0.16 <i>P_{RB} < .001 *</i>	0.24 ± 0.10 <i>P_{RB} < .001 *</i>	3.6 ± 1.6 <i>P_{RB} < .001 *</i>	7.6 ± 3.8 <i>P_{RB} < .001 *</i>
FBtoCB					
SICLE	0.74 ± 0.08 <i>P_{RT} < .001 *</i>	0.25 ± 0.12 <i>P_{RT} < .001 *</i>	0.27 ± 0.10 <i>P_{RT} < .001 *</i>	2.7 ± 1.1 <i>P_{RT} < .001 *</i>	6.0 ± 3.1 <i>P_{RT} < .001 *</i>
DEMONS	0.67 ± 0.12 <i>P_{RT} = .290, P_{SD} < .001 *</i>	0.23 ± 0.13 <i>P_{RT} < .001, P_{SD} = .005 *</i>	0.38 ± 0.17 <i>P_{RT} < .001, P_{SD} < .001 *</i>	4.2 ± 1.3 <i>P_{RT} < .009, P_{SD} < .001 *</i>	9.4 ± 3.2 <i>P_{RT} < .001, P_{SD} < .001 *</i>
RIGID-BONE	0.60 ± 0.14	0.30 ± 0.15	0.45 ± 0.16	4.1 ± 1.5	8.5 ± 3.5
RIGID-TUMOR	0.66 ± 0.12 <i>P_{RB} < .001 *</i>	0.24 ± 0.10 <i>P_{RB} < .001 *</i>	0.40 ± 0.16 <i>P_{RB} < .001 *</i>	3.6 ± 1.6 <i>P_{RB} < .001 *</i>	7.6 ± 3.8 <i>P_{RB} < .001 *</i>
CBtoFB					
SICLE	0.76 ± 0.08 <i>P_{RT} < .001 *, P_S < .001 *</i>	0.24 ± 0.10 <i>P_{RT} < .809 *, P_S = .490</i>	0.23 ± 0.12 <i>P_{RT} < .001 *, P_S = .197</i>	2.6 ± 1.0 <i>P_{RT} < .001 *, P_S < .001 *</i>	5.9 ± 3.1 <i>P_{RT} < .001 *, P_S = .015 *</i>
DEMONS	0.74 ± 0.10 <i>P_{RT} < .001 *, P_{SD} < .998 *, P_D < .001 *</i>	0.19 ± 0.10 <i>P_{RT} < .001 *, P_{SD} < .001 *, P_D = .152 *</i>	0.29 ± 0.14 <i>P_{RT} < .001, P_{SD} < .001 *</i>	3.3 ± 1.8 <i>P_{RT} < .001 *, P_{SD} < .001 *</i>	7.2 ± 4.6 <i>P_{RT} < .013 *, P_{SD} = .003 *</i> <i>P_D < .001 *</i>

Abbreviations: DSC; Dice similarity coefficient; FP = false positive; FN = false negative; ASD = average symmetric distance; FB = fan beam; CB = cone beam; D_{90%} = 90% confidence interval.

Statistical Tests: If the Shapiro-Wilk test for normality was passed, a Student paired *t*-test was used; if the Shapiro-Wilk test for normality was not passed, the Wilcoxon signed rank test was used (*). *P_{RB}* = comparison to RIGID-BONE; *P_{RT}* = comparison to RIGID-TUMOR; *P_{SD}* = comparison between SICLE and DEMONS; *P_S* = comparison between SICLE CBtoFB and FBtoCB; *P_D* = comparison between DEMONS CBtoFB and FBtoCB.

Table 2

Mean values of the evaluation measures for phase-to-phase registrations

Algorithm	DSC $\pm$ SD	
	CBCT	FBCT
RIGID BONE	0.75 ± 0.15	0.76 ± 0.12
RIGID TUMOR	0.78 ± 0.07 $P_{RB}=.055$	0.81 ± 0.09 $P_{RB}=.145$
SICLE	0.80 ± 0.07 $P_{RT}<.001^*$	0.82 ± 0.09 $P_{RT}=.305$
DEMONS	0.79 ± 0.07 $P_{RT}=.071^*$	0.81 ± 0.08 $P_{RT}=.876$

Abbreviations: DSC; Dice similarity coefficient; CBCT = cone beam CT; FBCT = fan beam CT.

The P_{RB} and P_{RT} values are based on a comparison of the result to the RIGID BONE and RIGID TUMOR registrations, respectively. If the Shapiro-Wilk test for normality was passed, a Student paired t -test was used; if the Shapiro-Wilk test for normality was not passed, the Wilcoxon signed rank test was used (*).