

Deformable Image Registration in MIM Maestro® Evaluation and Description

Jon Piper, Aaron Nelson, MD, and Jason Harper
MIM Software Inc.

Introduction

Deformable Image Registration (DIR) has found many uses in radiation therapy. It has been used as a time saving tool for generating entire structure sets for initial planning using atlas-based segmentation [1-6] or for adaptive therapy [7-11].

DIR has also been used to bring diagnostic images, not taken in the treatment planning position, into alignment with the treatment planning CT. DIR has been shown to provide more accurate registration of the tumor volume than rigid registration in this scenario [12,13]. Additionally, DIR is important for dose summation for tumor recurrence, adding dose between different treatment modalities, or tracking dose delivered during treatment.

Considering these uses of DIR, a robust, accurate, method of deformation is needed that can correct both for large changes in anatomy while also correcting for finer local differences. In response to this need, we have developed a constrained free-form intensity-based deformation method with very high degrees of freedom designed for same modality registrations. A description of the MIM® intensity-based deformation method and results from both phantom and patient studies are presented below.

Phantom and patient studies are important for characterizing an algorithm, however, it is even more important to be able to evaluate how the algorithm performed for a specific patient in the clinic (patient-specific QA). Once areas of error can be identified, then having a method to correct the registration in these areas is key. Reg Reveal® and Reg Refine™ (patent pending) have accomplished this goal. Results from these methods are also presented below.

Algorithm

The VoxAlign Deformation Engine® provides a constrained intensity-based, free-form deformable registration algorithm designed for same modality deformable registrations.

The deformation is initialized using a rigid registration that can be defined automatically or manually. Manual adjustment to the rigid registration allows the user to closely align a region of interest resulting in smaller movements during deformation. Regions that are most closely aligned will usually result in the most accurate deformation.

A course-to-fine multi-resolution approach is used to define a grid of control points on the static image which are used to search for the best, corresponding location in the target volume. Gross differences from the rigid registration are accounted for first and then refined to account for smaller local changes. The multi-resolution approach allows for good alignment even with large anatomical changes such as different arm or neck position and different phases of respiration. The final resolution of the computed deformation is no less than 3mm in each dimension.

The image matching metric minimizes intensity differences between the two images. The optimization strategy used is a custom modified gradient descent.

The deformation is regularized to avoid tears and folds in the deformation field. The regularization minimizes the effects of noise and incorrect correspondence while still providing a large degree of freedom for each control point to properly match the target volume.

Additional constraints are placed on the algorithm to further guide it towards reasonable results when executed on the same patient.

Caution should be employed when attempting to extrapolate performance of an algorithm based on other implementations, no matter how similar the strategy seems. Brock 2010 showed that such generalizations are difficult to make and that individual implementation can have a much greater impact on performance than algorithm characteristics [14]. Therefore it is essential to evaluate each DIR implementation on its own merit.

Evaluation Methods

Results have been presented for MIM's intensity-based deformable registration algorithm using real CT data [15,17], synthetic phantoms where deformations were applied to real CT data [15,16] and manually defined contours on real CT data [1-8].

Real CT data

Correlation: The deformable registration was evaluated for head-and-neck by comparing the correlation of a deformed CT with the target CT to the correlation after rigid registration. A reasonable benchmark was obtained by correlating the target with a shifted version of the target 1mm in each direction (resulting in 1.4mm total error) [15].

Landmark Accuracy: Deformable registration accuracy was assessed on the lung using the POPI model [17] where landmark positions were defined on two phases of a 4DCT and evaluated by measuring the displacement of a propagated landmark from a source image to a known landmark position in a target image. The 0% phase served as the source image and the 50% phase was the target image. One POPI model (patient1) displayed typical respiratory motion, 6.3mm +/- 3.0mm while the other (patient2) displayed significant respiratory motion, 14.0mm +/- 7.2mm, between end-expiration and end-inspiration. The 0% phase image was both directly registered to the 50% phase (DIRECT) and through intermediary phases (SERIAL) [19].

Deformable QA Evaluation: The ability to characterize deformable registration accuracy was also evaluated using Reg Reveal for the 0% to 50% phase deformable registration [20]. The evaluation method used was an interactive visual display of a rigid fusion between the

two volumes. This rigid fusion is computed as the best approximation to the local deformable registration, in a least squares sense, constrained such that the center of the rigid fusion is the exact transform defined by the deformable vector field at that point.

Two users independently reviewed the locally approximated rigid fusions centered at the points of interest. The users then rated their confidence using a binary scale that the approximated rigid registration represented an accurate alignment of the local anatomy within approximately 3 mm. These ratings were then compared to the measured error in the deformable registration and compared to each other.

Deformable Refinement: A method to influence the DIR algorithm to achieve a more accurate result (Reg Refine) was evaluated using patient2 from the POPI model and the DIRECT DIR method as a starting point [21]. The DIR was first evaluated with Reg Reveal to determine areas of the registration that needed improvement. Rigid registration adjustment tools were then used in areas where the naïve DIR was determined to be inaccurate to allow the observer to manually adjust the local registration or to execute an automatic rigid registration within a box of interest. The observer then recorded this preferable local alignment. When re-executing the DIR, these recorded local rigid alignments were used as inputs to influence the algorithm to achieve a local DIR closer to this observer-defined result.

Synthetic Phantom (deformed real CT data)

Consistency: A desirable property of a deformation algorithm is that it be consistent, that is, the concatenation of the forward and reverse deformation of an image pair should approximate the unity transform. Although, the algorithm has no constraints to encourage this behavior, the mean difference of the concatenated deformation field from the unity transform was calculated to evaluate the consistency of the algorithm using a head-and-neck CT [15].

Voxel Accuracy: Synthetic phantoms based on real clinical patient data were created by Nie et al 2012 for a prostate, head-and-neck, and prone/supine cranio-spinal case which extended from the base of the skull to

upper thighs. These phantoms provide the advantage of emulating the environment the algorithm will be used in clinically and therefore reflects the anatomical variations, detail, contrast, noise and artifacts found in real CT images. The test data was created using commercially available software, ImSimQA™, to create a realistic simulated deformation from each clinical patient case. Clinical images were used to guide the synthetic deformation and included bladder filling, soft tissue deformation, vertebral body movement, and tumor shrinkage/expansion. This process resulted in known deformations that were used to assess the deformable registration accuracy for MIM and two other commercially available algorithms [16].

Piper 2007 also created a synthetic phantom for head-and-neck by deforming a CT with a known deformation field, and the algorithm was evaluated for its ability to obtain this “gold standard” deformation after registration. The known deformation used was obtained when registering to the second CT for the test subject with another deformation algorithm [15].

Physical Phantom data

A single slice deformable prostate phantom was created by Kirby et al 2013 that represented a patient with an empty and full bladder [18]. Three distinct HU levels were created by mixing metallic additives with urethane rubber. Tissues types were homogeneous in intensity with noise introduced to provide heterogeneity. The single slice of the phantom was repeated numerous times to create a 3D image volume. Glow in the dark markers were applied to the phantom to track the deformed voxels. Voxel error and dice similarity indices were calculated.

Results – Head-and-Neck

Real CT data

Correlation [15]: The correlation coefficients after rigid registration were 0.890, 0.921, and 0.8593 for the three volume pairs. These improved to 0.979, 0.983, and 0.978, respectively, after deformable registration. These results were comparable to benchmark self-correlations of the target CTs with a 1.4mm translation error applied (0.979, 0.980, and 0.978, respectively).

Consistency [15]: Concatenating the forward and reverse

deformations resulted in an average distance of 3.1mm (3.1mm SD) from unity.

Synthetic Phantoms

Voxel Error: Voxel errors presented by Nie et al 2012 showed that MIM produced more accurate deformations in 97% of the voxels than the next best commercially available algorithm. Additionally, the percent of voxels with error greater than 1mm was 35% with MIM compared with 100% for the next best algorithm (Nie 2012, Fig. 3). The percent of voxels with error greater than 2mm was 15% with MIM compared to 70% for the next best algorithm (Nie 2012, Fig. 3).

Piper 2007 showed that Rigid registrations were unable to recover the “gold standard” displacements and resulted in an average of 10.3mm error (6.6mm SD). The deformable registration averaged 1.1mm error (1.9mm SD). Nearly three quarters (73.9%) of voxels had less than 1mm error and the 95% confidence interval was 4.8mm compared to 0.6% and 23.2mm respectively for rigid registration.

Contour Evaluation: Atlas-based segmentation studies have shown time savings ranging from 68-87% for normal structures and nodal targets [1]. Adaptive contouring studies has also shown an average time savings of 75% for normal structures and targets [8].

Results – Lung

Real CT data

Landmark Accuracy: For patient1, the residual errors after DIRECT and SERIAL deformation were 0.8mm +/- 0.4mm and 1.8mm +/- 1.4mm, respectively. For patient2, the errors after DIRECT and SERIAL were 5.1mm +/- 7.5mm and 1.6mm +/- 1.6mm, respectively. The mean residual errors for points of interest with <5mm, 5-10mm, 10-15mm, 15-20mm, 20-25mm, and 25+mm initial displacement were 0.9mm, 0.9mm, 1.4mm, 4.7mm, 9.7mm, and 20.2mm after DIRECT deformation and 1.3mm, 1.4mm, 2.1mm, 2.3mm, 1.9mm, and 3.1mm after SERIAL deformation, respectively.

Deformable QA Evaluation: Using the DIR between the 0% to 50% phases for patient2 of the POPI model, the two observers separated “good” and “concerning”

registrations with 91% and 92% accuracy at a threshold of 3.1mm and 2.4mm respectively (t-test, $p < 0.0000001$ for reach observer). Sensitivity for detecting errors greater than 5mm was 86% and 97% respectively. There also was good agreement between the two observers, with a kappa of 0.72.

Deformable Refinement: The DIRECT DIR method between the 0% to 50% phases for patient2 resulted in residual registration errors ranging from 0.3mm to 35.4mm (5.1 +/- 7.5mm). After recording seven local rigid registrations using Reg Refine and re-executing the DIR algorithm with these as inputs, the distribution of errors decreased to 0.2mm to 5.6mm (1.2 +/- 0.9mm), which was a significantly better result ($p < 0.000001$).

Contour Evaluation: Atlas-based segmentation studies have shown the accuracy of multi-atlas segmentation with an overall average dice similarity index of 0.81 for the contours tested: right and left lung, trachea, heart, and esophagus [6].

Results – Prostate

Physical Phantom

Voxel Error [18]: The mean error for MIM with default settings was 1.5mm with 9.8% of voxels having >3mm of error and 3.7% of voxels having > 7mm of error. The average dice similarity index for MIM was 0.96 which was the highest of the algorithms tested. With an increase of the regularization (smoothness) factor, the mean error was 1.3mm with 6.5% of voxels having >3mm error and 2.2% of voxels having >7mm error.

Synthetic Phantom (deformed real CT data)

Voxel Error: Voxel errors presented by Nie et al 2012 showed that MIM produced more accurate deformations in 99% of the voxels than the next best commercially available algorithm. Additionally, the percent of voxels with error greater than 1mm were 7% with MIM compared with 100% for the next best algorithm (Nie 2012, Fig. 3). The percent of voxels with error greater than 2mm were 2% with MIM, compared 30% for the next best algorithm (Nie 2012, Fig. 3).

Results – Craniospinal

Synthetic Phantom

Voxel Error: Voxel errors presented by Nie et al 2012 showed that MIM produced more accurate deformations in 56% of the voxels than the next best commercially available algorithm. The percent of voxels with error greater than 1mm were 40% with MIM compared with 50% for the next best algorithm (Nie 2012, Fig. 3). The percent of voxels with error greater than 2mm were 20% with MIM, compared 12% for the other algorithm (Nie 2012, Fig. 3).

Need for Patient-specific Evaluations

The aforementioned DIR accuracy phantom studies are only able to characterize the performance of the algorithm in synthetic or specific scenarios and will not necessarily reflect the accuracy of an algorithm on a particular patient. This, however, is the most important question in DIR. Every patient is different, and evaluating the accuracy of DIR can be critical when performing dose accumulation or diagnostic image DIR to planning images.

Reg Reveal was developed for this purpose and is currently the only tool available specifically for the purpose of efficient quality assurance of DIR. Reg Reveal allows the user to interrogate the registration in specific regions of interest and draw conclusions about it's accuracy. Such a deformable registration evaluation tool, which allows the user to answer the essential question of whether a DIR algorithm correctly identifies corresponding anatomy for a given patient, should be considered an essential QA tool when implementing DIR in the clinic. After knowing this information, Reg Refine allows the user to influence the registration algorithm to achieve a more accurate result.

References

1. Hu K, Lin A, Young A, Kubicek G, Piper JW, Nelson AS, Dolan J, Masino R, Machtay M. Timesavings for Contour Generation in Head and Neck IMRT: Multi-institutional Experience with an Atlas-based Segmentation Method. *IJROBP* 2008; 72(1):S391..
2. Pirozzi SD, Nelson AS, Piper JW. Atlas-based Segmentation: Comparison of Multiple Segmentation Approaches for Lymph Level Targets and Normal Structures in Head and Neck Cancer. *IJROBP* 2011; 81(2):S828.
3. Lin A, Kubicek G, Piper JW, Nelson AS, Dicker AP, Valicenti RK. Atlas-Based Segmentation in Prostate IMRT: Timesavings in the Clinical Workflow. *IJROBP* 2008; 72(1): S328-329.
4. Pirozzi SD, Horvat M, Nelson AS, Piper JW. Atlas-based Segmentation: Evaluation of a Multi-Atlas Approach for Prostate Cancer. *IJROBP* 2012; 84(3):S799.
5. Young AV, Wortham A, Wernick I, et al. Atlas-based segmentation improves consistency and decreases time required for contouring postoperative endometrial cancer nodal volumes. *IJROBP* 2011; 79 (3):943-947.
6. Pirozzi S, Horvat M, Piper J, Nelson AS. Atlas-based Segmentation: Evaluation of a Multi-Atlas Approach for Lung Cancer. *Med Phys* 2012; 39:3677.
7. Choe J, Piper JW, Nelson AS, Nelson AD, Kuc N, Kulasekera R, Greskovich J. Comparison of Contour-to-CT Registration Techniques During Adaptive Re-planning of Radiation Therapy for Patients with Bulky Disease. *IJROBP*. 2007;69(3 Suppl S):S725.
8. Fragoso RC, Piper JW, Nelson AS, Harrison AS, Machtay M, Xiao Y. Evaluation of a Deformable Re-Contouring Method for Adaptive Therapy. Presented at ACRO Annual Meeting 2008.
9. Tsuji SY, Hwang A, Weinberg V, et al. Dosimetric evaluation of automatic segmentation for adaptive IMRT for head-and-neck cancer. *IJBROBP* 2010; 77(3):707-714.
10. Ravi A, Smith G, Lee J, et al. An evaluation of geometric and dosimetric accuracy of using automatic deformable registration to improve the efficiency in adaptive radiation therapy treatment planning for head and neck cancer patients. *IJROBP* 2011; 81(2):S838-839.
11. Liu H, Greskovich J, Koefman S, et al. Evaluation of volumetric change and dosimetric discrepancy with daily cone-beam CT for patients with head-and-neck cancer. *Med Phys* 2012; 39:3782.
12. Kovalchuk N, Niedermayr TR, Oureshi MM, et al. Comparison of deformable versus rigid PET/CT registration in head-and-neck cancer patients undergoing postoperative radiation therapy. *IJROBP* 2012; 84(3):S511-512.
13. Fogh SE, Kubicek GJ, Axelrod R, et al. Utility of deformable PET fusion for gross tumor volumes in head & neck malignancies. Presented at ASTRO Translational Advances in Radiation Oncology and Cancer Imaging 2008.
14. Brock KK. Results of a Multi-Institutional Deformable Registration Accuracy Study (MIDRAS). *IJROBP* 2010; 76(2):583-596.
15. Piper JW. Evaluation of an intensity-based free-form deformable registration algorithm. *Med Phys* 2007; 34(6):2353-2354.
16. Nie K, Kirby N, Pouliot J, et al. Development of a workflow to quantitatively validate various image fusion algorithms in radiation therapy. Presented at ASTRO Annual Meeting 2012 – Poster 3384.
17. Vandemeulebroucke J, Rit S, Kybic J, et al. Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. *Med Phys* 2011; 38(1):166-178.
18. Kirby N, Chuang C, Ueda U, et al. The need for application-based deformable image registration. *Med Phys* 2013; 40(1):1-10.
19. Piper JW, Duchateau M, Nelson AS, et al. Characterizing errors in 4DCT deformable registration with the POPI Phantom. Submitted for presentation to AAPM 2013.
20. Piper JW, Nelson AS, Duchateau M, et al. A novel framework for deformable registration evaluation and quality assurance. Submitted for presentation to ASTRO 2013.
21. Piper JW, Nelson AS, Duchateau M et al. A novel framework for user-intervened correction of deformable registration. Submitted for presentation to ASTRO 2013.