HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2015 April 22.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2015 March 20; 9413: . doi:10.1117/12.2081045.

Evaluation of Five Image Registration Tools for Abdominal CT: Pitfalls and Opportunities with Soft Anatomy

Christopher P. Lee[a,*], **Zhoubing Xu**[b], **Ryan P. Burke**[c], **Rebeccah B. Baucom**[d], **Benjamin K. Poulose**[d], **Richard G. Abramson**[e], and **Bennett A. Landman**[a,b,c,e]

[a]Computer Science, Vanderbilt University, Nashville, TN, USA 37235

[b]Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235

[c]Biomedical Engineering, Vanderbilt University, Nashville, TN, USA 37235

[d]General Surgery, Vanderbilt University, Nashville, TN, USA 37235

[e]Radiology and Radiological Science, Vanderbilt University, Nashville, TN, USA 37235

Abstract

Image registration has become an essential image processing technique to compare data across time and individuals. With the successes in volumetric brain registration, general-purpose software tools are beginning to be applied to abdominal computed tomography (CT) scans. Herein, we evaluate five current tools for registering clinically acquired abdominal CT scans. Twelve abdominal organs were labeled on a set of 20 atlases to enable assessment of correspondence. The 20 atlases were pairwise registered based on only intensity information with five registration tools (affine IRTK, FNIRT, Non-Rigid IRTK, NiftyReg, and ANTs). Following the brain literature, the Dice similarity coefficient (DSC), mean surface distance, and Hausdorff distance were calculated on the registered organs individually. However, interpretation was confounded due to a significant proportion of outliers. Examining the retrospectively selected top 1 and 5 atlases for each target revealed that there was a substantive performance difference between methods. To further our understanding, we constructed majority vote segmentation with the top 5 DSC values for each organ and target. The results illustrated a median improvement of 85% in DSC between the raw results and majority vote. These experiments show that some images may be well registered to some targets using the available software tools, but there is significant room for improvement and reveals the need for innovation and research in the field of registration in abdominal CTs. If image registration is to be used for local interpretation of abdominal CT, great care must be taken to account for outliers (e.g., atlas selection in statistical fusion).

Keywords

Image Registration; Computed Tomography; Abdomen

[*]christopher.p.lee@vanderbilt.edu; http://masi.vuse.vanderbilt.edu; Medical-image Analysis and Statistical Interpretation Laboratory, Department of Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235.

1. INTRODUCTION

The segmentation of the abdomen is extremely important for clinical analysis and medical engagement. Manual labeling has been the favored approach for producing trustworthy segmentations, but often is burdened with unreasonable with time and resource constraints. In large-scale studies, robust automatic abdominal segmentation becomes necessary. Atlas-based segmentation provides a non-parametric solution by transferring existing segmentations on standard atlases to the target image through registration, where the quality of inter-subject registrations has been the crux of this type of approaches.

General-purpose registration tools from volumetric brain registration and are now being applied to abdominal computed tomography (CT) scans. Compared to the relatively consistent brain anatomy, human abdomens present a huge number of variations that complicates the registrations. Besides the inter-subject differences (e.g., age, gender, stature, normal anatomical variants, disease status), soft anatomy within the abdomen deforms vastly within individuals (e.g., pose, respiratory cycle). While more substantial errors can be expected, caveats should be taken for atlas-based abdominal segmentation in the context of non-robust abdominal CT registrations. This prompts the need for the performance evaluation of existing registration tools on abdominal CTs, with a special focus on this application atlas-based segmentation.

Previously, Klein et al. [1] applied 14 nonlinear registration tools and one linear registration algorithm on MRIs of the human brain to identify the nonlinear deformation algorithms most tailored for brain image registration. In their study, the registrations were evaluated based on the Valmet validation tool, where 3-D object segmentations were assessed using both volume- and surface-based metric criteria [2].

In this study we assessed four registration tools that have been successful in volumetric brain registrations, including FNIRT [3], IRTK [4], NiftyReg [5], and ANTs [6], due to their academic popularity and immediate availability. A common affine registration procedure (using the rigid and affine registration of IRTK [7]) was conducted first as the baseline of the following non-rigid registrations using the four registration tools. The efficacies of the non-rigid registration algorithms using the four registration tools based on a common starting point of affine registration (using rigid and affine registration tools of IRTK) were evaluated based on Dice similarity coefficient (DSC), mean surface distance (MSD), and Hausdorff distance (HD).

2. METHODS

2.1 Data

Twenty abdominal CT scans were randomly selected from an ongoing colorectal cancer chemotherapy trial under Institutional Review Board (IRB) supervision in anonymous form and acquired in NIFTI format. To reduce regions of confusion, all 20 scans were first cropped along the cranio-caudal axis to include liver, spleen, and kidneys entirely with a tight border. After the cropping, variable field of views (approx. 300 × 300 × 200 mm ~ 500 × 500 × 300 mm) and resolutions (approx. 0.6 × 0.6 × 3.0 mm ~ 1.0 × 1.0 × 5.0 mm) were

captured. Before any further processing, the image orientations were normalized in the NIFTI header. Twelve abdominal organs were considered as regions of interest (ROI), including spleen, right kidney, left kidney, gall bladder, esophagus, liver, stomach, aorta, inferior vena cava, portal & splenic vein, pancreas, and the adrenal glands. These twelve ROIs were manually labeled by two experienced undergraduate students, and then verified by a radiologist to enable assessment of correspondence. Inter-rater agreement was evaluate on a subset of 6 datasets, which were completely and independently labeled twice. Mean DSC overlap between the raters was 0.86 ± 0.03.

2.2 Basic Registration Pipeline

As illustrated in Figure 2, for each target image among the 20 scans, the remaining 19 atlases were used as source images to the target image in a pair-wise manner, thus 380 sets of output were generated. For each pair of the atlas and target, the registration was first driven by the dissimilarity metrics between their intensity images. The associated atlas label was then propagated to the target space with nearest neighbor interpolation as the estimate of the target structures based upon the transformation / deformation generated from the intensity-driven registration. In the end, the registered labels on the twelve ROIs were validated against the manual segmentation as the evaluation of the registration results. For all software packages, we used the default parameters of the registration tools except as noted below..

2.3 Specific Registration Setups

2.3.1 Affine Registration—The rigid and affine registration tools of the IRTK toolkit were selected to yield the affine registrations. A rigid registration was first applied, and the affine registration was conducted with the initialization of the rigid transformation. For both procedures, the target padding value was set as −900 to reduce the impact of the background in the scan, considering that the Hounsfield unit of air was −1024. These two linear registrations were also conducted using a coarse-to-fine scheme on three resolution levels. Assuming relatively homogenous orientations of patients' bodies in the CT scans, we specified the options of "translation_only" and "translation_scale" for the rigid and affine registration, respectively, so that only translation (and scaling for the affine registration) adjustments were allowed, and the searches over rotations were prohibited. We used the affine registration results as the starting point (source image) for the following non-rigid registrations.

2.3.2 Nonlinear Registrations—For FNIRT, we followed the default parameter specifications.

The IRTK non-rigid registration used the same coarse-to-fine scheme and target padding value as its linear counterparts. Furthermore, we specified the B-spline control point spacing to be 20, 10 and 5mm for three stages of the non-rigid registration.

For the Nifty registration, we specified the weight of the bending energy penalty term to be 0.00001, and the grid spacing as five voxels for each orientation.

The ANTs non-rigid registration used a five level multi-resolution sampling. For the five levels' registrations, the shrink factors were 10, 6, 4, 2, and 1, the smoothing factors were 5, 3, 2, 1, and 0 voxels, and the numbers of iterations were 100, 100, 70, 50, and 20, respectively. We specified the cross-correlation as the image dissimilarity metric with the neighborhood radius of 4. The window of the intensity values considered was set within the range of [0.005, 0.995]. The symmetric normalization (SyN) transform was used with the gradient step as 0.1, update field variance as 3, and total field variance as 0.

2.4 Quantitative Validation

Using the manually segmented labels as truth values for the target images, and the registered source labels from each nonlinear registration tool respectively, metrics were calculated. To begin with, DSC, MSD, and HD were calculated on the registered organs individually. Next, as expected of lots of registration failures, we retrospectively selected the top 1 and 5 atlases for each target image and calculated the corresponding DSC values. To further our understanding, we constructed majority vote segmentation with the top 5 DSC values for each organ and atlas.

3. RESULTS AND DISCUSSION

Four of the registration tools produced 380 pairwise registered labels. FNIRT had two atlases that failed without producing output and only produced 378. We first established the metrics on each of the registration tool's raw data using DSC, MSD, and HD on each organ, by comparing the registered labels with the manually segmented labels as the ground truth. In Figure 3, as demonstrated by the DSC values for each organ, it becomes immediately apparent that the smaller organs have much smaller values and the larger organs have larger values. This is surely caused by the fact that larger organs have more surface area, and therefore have a greater probability of overlap. However, the DSC results display slight superiority in registration accuracies with non-rigid IRTK, NiftyReg, and ANTs, and inferiority with affine IRTK and FNIRT. The MSD and HD boxplots clearly illustrate the overbearing amount of outliers in our raw data set. The MSD outliers reach just less than 300mm and the HD outliers reach over 300mm. On the MSD and HD boxplots, the dominance of any registration tool is indistinguishable. Due to each registration creating outliers and skewing the raw results in Figure 3, it became imperative to evaluate the results that were not catastrophic failures.

To evaluate the better-performing results, we averaged the DSC values of all the organs for each atlas and selected the top 1 and 5 atlases as shown in Figure 4. By averaging the organs, we can understand how the registration tools are performing in general for "successful" cases that may be selected by an atlas selection procedure. By using retrospective analysis, the results clearly serve as an upper bound on average registration performances considering only "successes."

The cleaner interpretation of the results in Figure 4 fortifies the conclusion drawn on the raw DSC results; non-rigid IRTK, Nifty and ANTs perform better than affine IRTK and FNIRT.

To further our understanding, we constructed majority vote segmentation by selecting the top 5 DSC values for each organ and target image (Figure 5). In figure 5, the DSC of the majority vote segmentation furthers the argument that affine IRTK and FNIRT perform less accurate registrations than non-rigid, IRTK, NiftyReg, and ANTs. For the most part, IRTK achieves better registrations on the larger organs and NiftyReg produces better registrations on the smaller organs. Comparing the DSC of the raw results in Figure 3 to the DSC of the majority vote segmentation in Figure 5, the results illustrated a median improvement of 85% through the majority vote segmentation.

4. CONCLUSION

In this paper, we analyze 5 different general-purpose image registration tools and apply them to abdominal CT scans. We use one linear registration tool, affine IRTK, and four non-linear registration tools, FNIRT, NiftyReg, non-rigid IRTK, and ANTs. In general, we used the default parameters in our registrations and believe that the registrations may improve with further inspection on parameter selection. However, the results of our experiment outline that affine IRTK and FNIRT produce worse registrations than non-rigid IRTK, NiftyReg and ANTs. Furthermore, it became apparent that every registration tool produced catastrophic failures. These catastrophic failures detracted from the well-registered atlases. When applying these image registrations tools for local interpretation of abdominal CT scans, great care must be taken to account for outlier (e.g., atlas selection in statistical fusion). With the current tools, registrations of abdominal CTs are too erroneous and sporadic to be implemented in clinical use. Development and innovation in the field of image registration for abdominal CT scans are critical to reduce outliers, maintain registration consistency and improve organ identification and segmentation.

ACKNOWLEDGMENTS

This research was supported by NIH 1R03EB012461, NIH 2R01EB006136, NIH R01EB006193, ViSE/VICTR VR3029, NIH UL1 RR024975-01, NIH UL1 TR000445-06, NIH P30 CA068485, and AUR GE Radiology Research Academic Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

1. Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage. 2009; 46(3):786–802. [PubMed: 19195496]
2. Gerig G, Jomier M, Chakos M. Valmet: A new validation tool for assessing and improving 3D object segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI. 2001; 2001:516–523.
3. Jenkinson M, Beckmann CF, Behrens TE, et al. Fsl. NeuroImage. 2012; 62(2):782–790. [PubMed: 21979382]
4. Rueckert D, Sonoda LI, Hayes C, et al. Nonrigid registration using free-form deformations: application to breast MR images. Medical Imaging, IEEE Transactions on. 1999; 18(8):712–721.
5. Modat M, Ridgway GR, Taylor ZA, et al. Fast free-form deformation using graphics processing units. Computer methods and programs in biomedicine. 2010; 98(3):278–284. [PubMed: 19818524]
6. Avants BB, Epstein CL, Grossman M, et al. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis. 2008; 12(1):26–41. [PubMed: 17659998]

7. Studholme C, Hill DL, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. Pattern recognition. 1999; 32(1):71–86.

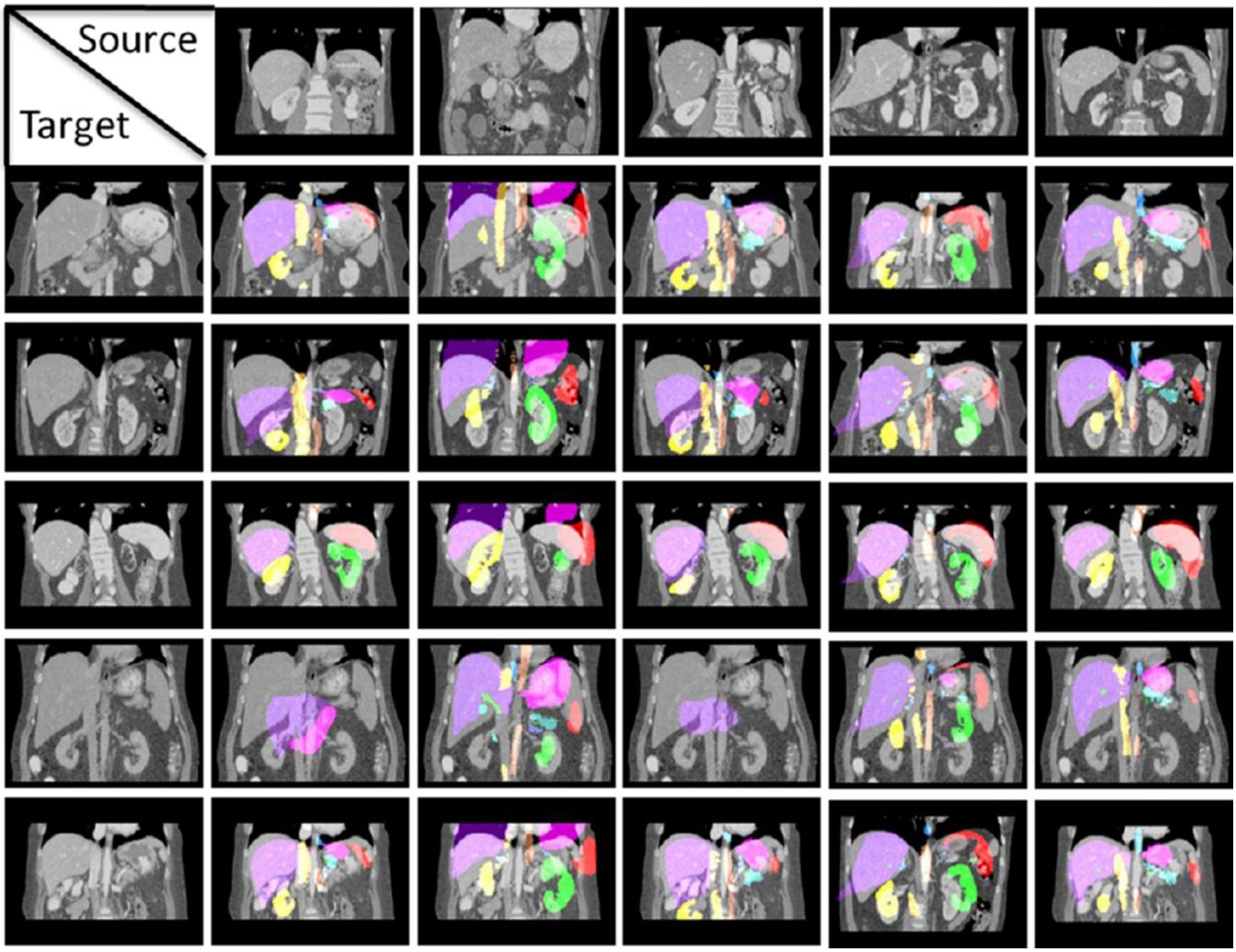

Figure 1.
Examples illustrate the variability of image registrations in abdominal CT.

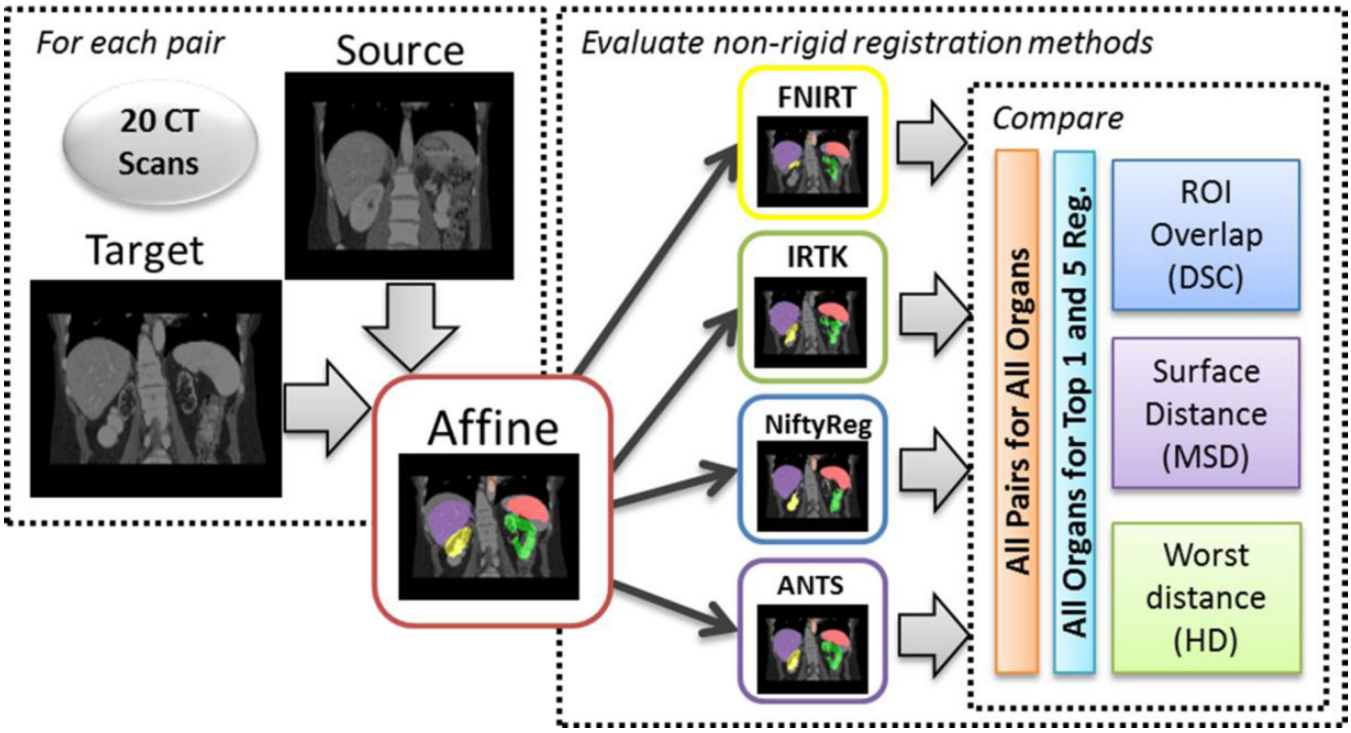

Figure 2.

The proposed general pipeline for conducting registrations and followed by metric analyses. Initially, each of 20 CT scans were pairwise linear registered using affine IRTK. Using the linear registration as a baseline, the registrations went through four non-rigid registrations. The output non-rigid registrations are evaluated against the manual segmentation via the DSC overlap, mean surface distance, and Hausdorff distance for each organ of interest. The comparison of the four non-rigid registrations was first based on all pairs of inter-subject registrations for all organs. Then another round of comparison was applied to the top 1 and top 5 registrations selected retrospectively for each target CT scan.

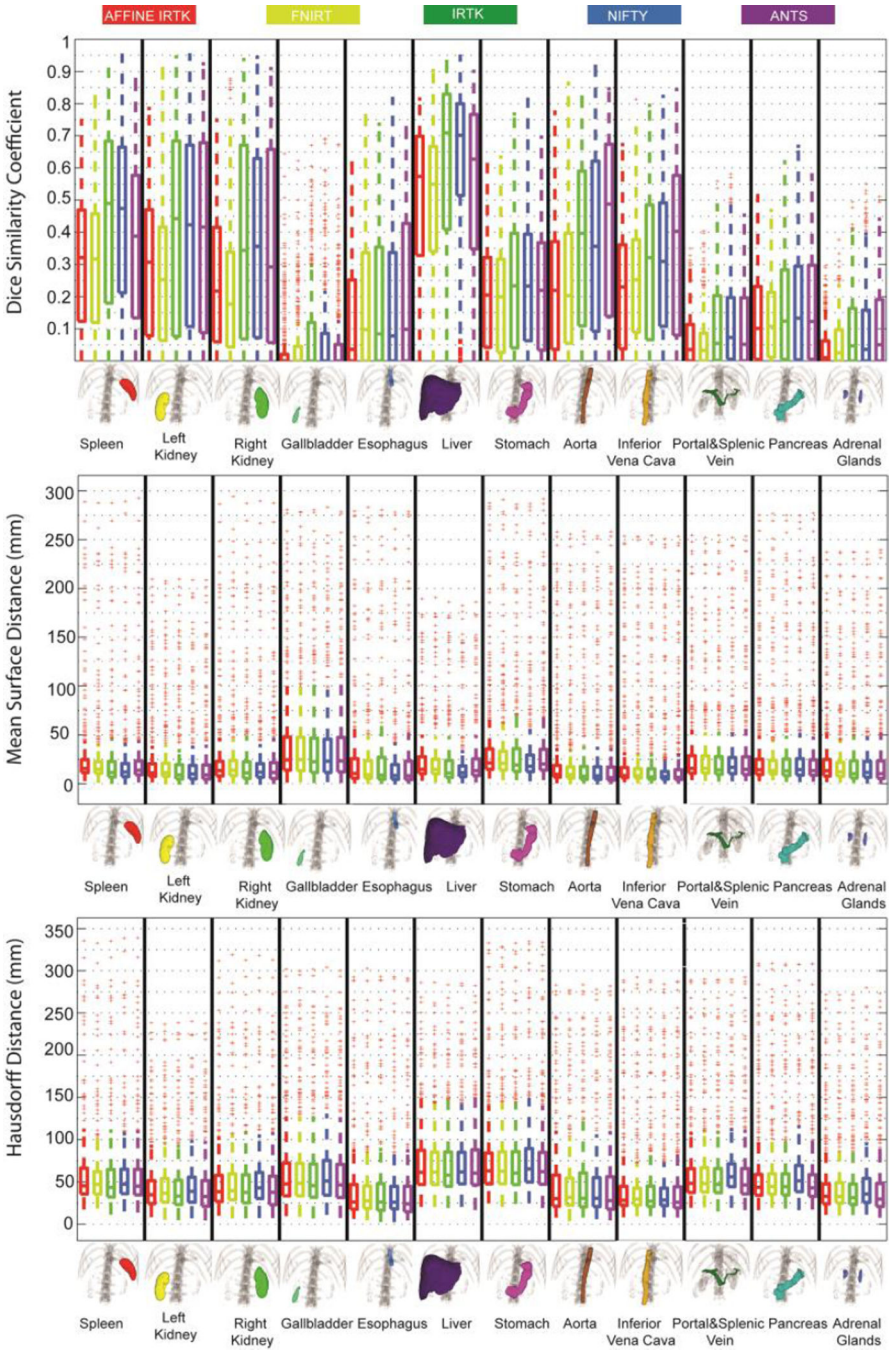

Figure 3.

The quantitative comparisons of one affine registration and four non-rigid registrations on 12 regions of interest in terms of DSC, MSD, and HD, respectively.

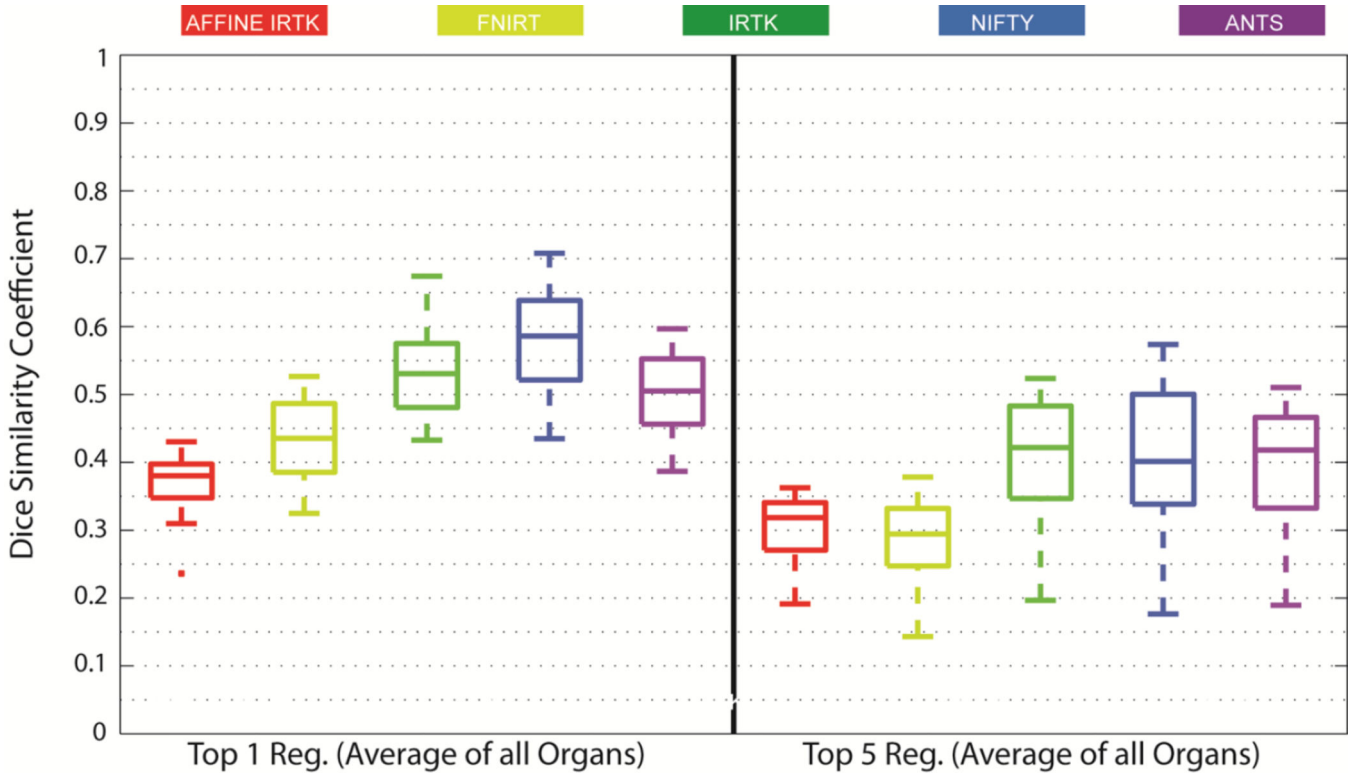

Figure 4.

Quantitative results comparing one affine registration and four non-rigid registrations based on the top 1 and 5 registrations (retrospectively selected) for 2o target CT scans using the average DSC over all organs.

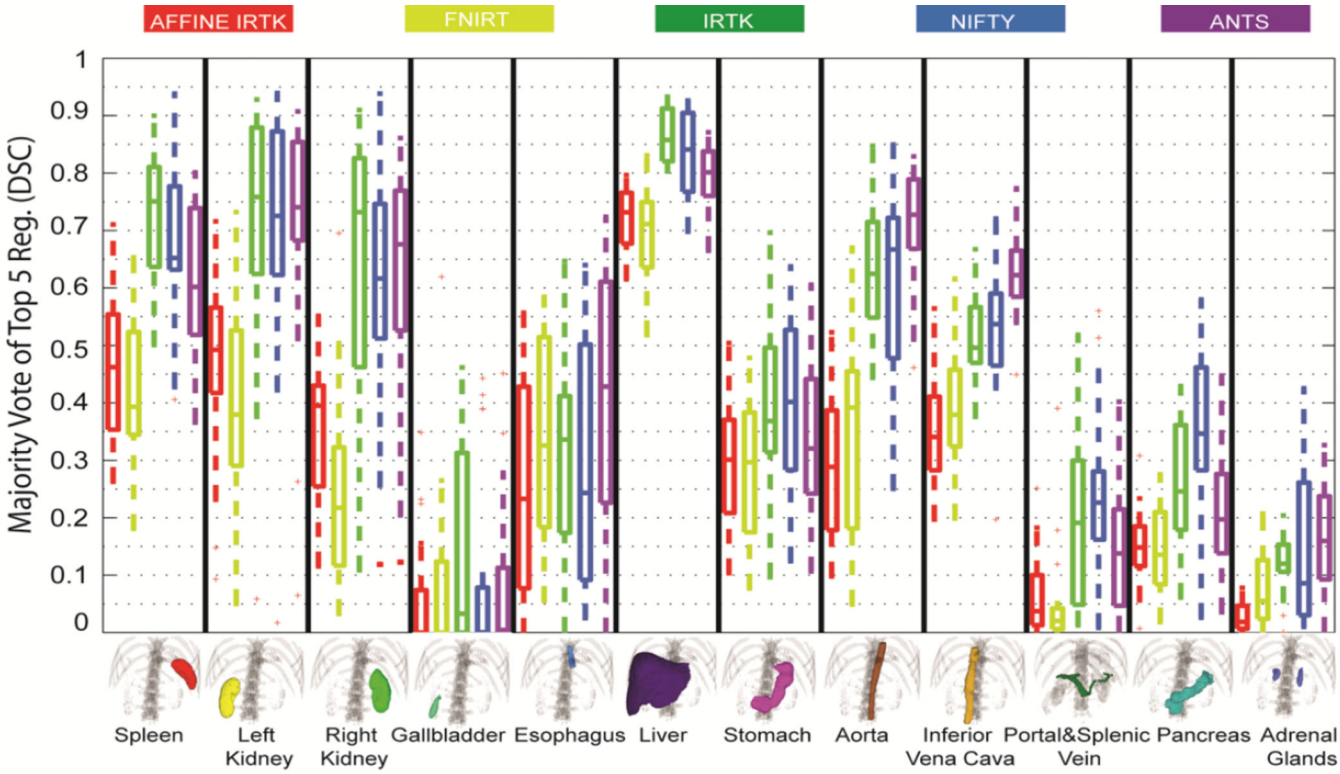

Figure 5.

Quantitative results comparing the MV fusion of the registrations with the top 5 DSC values for each organ on 20 target CT scans. The DSC values of the MV fusion were used as the criteria to compare the baseline multi-atlas segmentation for the tested registration tools.