Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Automatic selection of landmarks in T1-weighted head MRI with regression forests for image registration initialization

Jianing Wang
Yuan Liu
Jack H. Noble
Benoit M. Dawant

SPIE.

Jianing Wang, Yuan Liu, Jack H. Noble, Benoit M. Dawant, "Automatic selection of landmarks in T1-weighted head MRI with regression forests for image registration initialization," *J. Med. Imag.* **4**(4), 044005 (2017), doi: 10.1117/1.JMI.4.4.044005.

Automatic selection of landmarks in T1-weighted head MRI with regression forests for image registration initialization

Jianing Wang,* Yuan Liu, Jack H. Noble, and Benoit M. Dawant
Vanderbilt University, Vanderbilt Institute for Surgery and Engineering, Department of Electrical Engineering and Computer Science, Nashville, Tennessee, United States

Abstract. Medical image registration establishes a correspondence between images of biological structures, and it is at the core of many applications. Commonly used deformable image registration methods depend on a good preregistration initialization. We develop a learning-based method to automatically find a set of robust landmarks in three-dimensional MR image volumes of the head. These landmarks are then used to compute a thin plate spline-based initialization transformation. The process involves two steps: (1) identifying a set of landmarks that can be reliably localized in the images and (2) selecting among them the subset that leads to a good initial transformation. To validate our method, we use it to initialize five well-established deformable registration algorithms that are subsequently used to register an atlas to MR images of the head. We compare our proposed initialization method with a standard approach that involves estimating an affine transformation with an intensity-based approach. We show that for all five registration algorithms the final registration results are statistically better when they are initialized with the method that we propose than when a standard approach is used. The technique that we propose is generic and could be used to initialize nonrigid registration algorithms for other applications. [DOI: 10.1117/1.JMI.4.4.044005]

Keywords: image preregistration initialization; landmark selection; regression forest; random sample consensus.

Paper JMI-17107R received Feb. 3, 2017; accepted for publication Sep. 27, 2017; published online Nov. 14, 2017.

1 Introduction

Medical image registration establishes a correspondence between images of biological structures, and it is at the core of many applications. Most deformable registration methods that are commonly used are dependent on a good preregistration initialization.[1,2] The initialization can be performed by manually aligning the images, localizing homologous landmarks, and calculating a point-based transformation between the images or with intensity-based affine registration techniques.

When landmarks are used, the selection of these landmarks is important. Good landmarks should cover the entire biological structure and should be easy to localize unequivocally, i.e., they should have distinct and salient features. While manual selection of landmarks is possible for small landmark sets, it becomes impractical for larger sets that are required to, for instance, register nonrigidly three-dimensional (3-D) image volumes. In this paper, we propose a learning-based method to automatically find a set of robust landmarks in 3-D MR image volumes of the head to initialize nonrigid transformations. Our methods involve two steps. First, landmarks that can be reliably localized in the images are identified using a random forest (RF)-based method.[3] The subset of landmarks that leads to good initialization transformations, which are computed with a thin plate spline-based (TPS) method,[4] is then selected using a random sample consensus (RANSAC) algorithm.[5]

To show the value of our registration initialization technique, we compare the final registration results obtained with five well-established deformable registration algorithms, i.e., (1) Adaptive Bases Algorithm (ABA),[6] (2) Automatic Registration Tools (ART),[7] (3) Diffeomorphic Demons (DD),[8] (4) Fast Free Form Deformation (F3D),[9] and (5) Symmetric Normalization (SyN),[10] when either an affine transformation or the proposed method is used for preregistration initialization. We show that a higher registration accuracy is achieved in the latter case.

2 Methods

2.1 *Overview*

Our dataset contains images of 201 individuals from the data repository we have created over a decade for deep brain stimulation surgeries.[11] All of these images are T1-weighted sagittal MR image volumes with approximately $256 \times 256 \times 170$ 1 mm^3 isotropic voxels. All have been acquired clinically with the subjects in roughly the same position but without special care being taken to position them. The images are randomly partitioned into a first training dataset of 100 images that is used to train RF models[3] that are used to localize a set of landmarks, a second training dataset of 80 images that is used to identify among the set of landmarks which ones are the most robust, a testing dataset of 20 images that is used to validate our technique, and one atlas image.

Our technique includes four training steps: (1) the generation of the candidate landmark set using the atlas image, (2) the creation of a series of RF models that are each trained to localize one landmark, (3) the localization of the candidate landmarks in

*Address all correspondence to: Jianing Wang, E-mail: jianing.wang@vanderbilt.edu

Table 1 Comparison of the five deformable registration algorithms.

Algorithm	Deformation model	Similarity measure	Regularization
ABA	Radial basis functions	NMI of whole brain	Transformation symmetry; Jacobian threshold
ART	Homeomorphism	NCC of whole brain	Gaussian smoothing
DD	Diffeomorphic optical flow	SSD of whole brain	Gaussian smoothing
F3D	Cubic B-splines	NMI of whole brain	Bending energy
SyN	Symmetric diffeomorphism	CC of whole brain	Gaussian smoothing; transformation symmetry

NMI, normalized mutual information; SSD, sum of square differences; CC, cross correlation; NCC, normalized cross correlation.

the second training dataset, and (4) the selection of the most reliable landmarks using a RANSAC algorithm and the second training set. In the testing phase, the most reliable landmarks are localized in unknown volumes and they are used to compute a smoothing TPS transformation that registers the atlas to each of the test volumes. Further refinement of the initial registration is performed with the five deformable registration algorithms. Differences between these algorithms are briefly summarized in Table 1.[12] Each registration algorithm requires values for a set of parameters, and we use the method presented by Liu et al.[12] to select them.

2.2 *Generation of the Candidate Landmark Set*

The brain in the atlas image volume is extracted with the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain Software Library (FSL) brain extraction tool.[13] Five thousand candidate landmarks are randomly placed inside the brain region (Fig. 1). To find the position of the candidate landmarks on the 180 training images, the atlas image is first registered to each of the training images using a sequence of intensity-based rigid and nonrigid registration steps.[6] The accuracy of the registration is visually assessed, and each of the landmarks is projected from the atlas image to each of the training images. These projected landmarks are considered to be the ground truth position of the landmarks in the training images, and we denote the true position of landmark L_i in image I_j as $T_{i,j}$.

2.3 *Creation of the Random Forest Models*

To reduce computation cost, we downsample the training images in the first training dataset by a factor of 4. An RF model is trained to localize each landmark in the downsampled images using the methods presented by Glocker et al.[14] Briefly, given a point in a training image, a set of multiscale long-range textural features are extracted and associated to a probability that this point is at the position of the landmark.[15] The RF model is trained to learn the relationship between the features and the

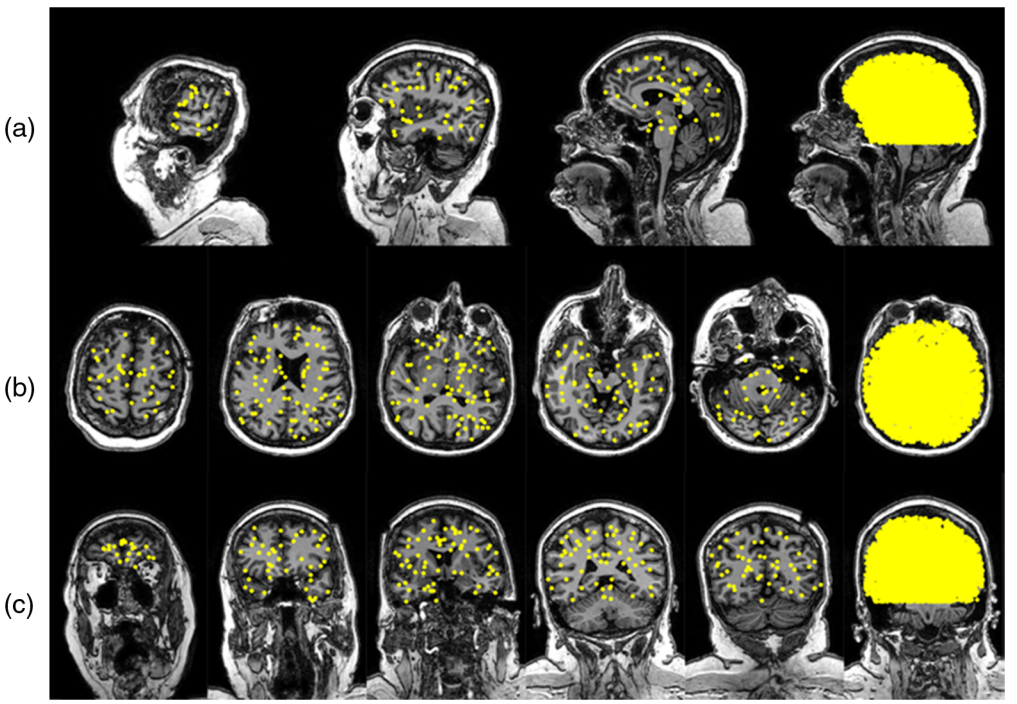

Fig. 1 Candidate landmarks shown on selected slices of the atlas image volume in the sagittal (row 1), axial (row 2), and coronal (row 3) directions. The last image of each row is created by projecting all landmarks on the same slice to show the region of the head covered by the landmarks.

probability value. Because one RF model is trained for each landmark, 5000 RF models are trained.

2.4 *Localization of the Candidate Landmarks*

The images in the second training dataset are first downsampled by a factor of 4. The RF models trained in the previous step are then applied to localize the candidate landmarks in these downsampled images. Given an unknown image and one RF model that is trained to localize a specific landmark L_i, the output of the RF model is a 3-D probability map of the same size as the input image (Fig. 2). In this map, a high value indicates a high probability that the point is the landmark of interest. The local maxima of the 3-D probability map are thus the potential positions of the landmark in the image. Because the landmark should be located inside the brain region, the local maxima that are close to the border of the 3-D probability map are likely to be false positives. Such local maxima are discarded, and, if no local maximum is left, then we consider that the landmark is not localizable in the input image.

We trim the candidate landmark set by removing landmarks that are not localizable in all of the 80 images in the second training dataset. Thus, each of the remaining candidate landmarks has at least one and possibly several corresponding potential landmark points in each training image. Possible reasons for multiple localization include the lack of salient features around a point or anatomic variations. In this work, we select only one point from the multiple potential landmark points. To do so, we define n as the number of images among the 80 images in the second training set in which a landmark has only one corresponding landmark. If $n < 30$, we consider this landmark to be hard to localize unequivocally and we remove it from the candidate landmark set. If $n \geq 30$, then the mean position of the landmarks in these n images is calculated, and, for each of the 80-n images, the point that is closest to the mean is selected as the landmark in this image. Because the landmark is localized in the downsampled image, we calculate a coarse landmark position in the full resolution image by upsampling the coordinates of the landmark, and we denote the coarse position of landmark L_i in the full resolution image I_j as $R_{i,j}$. If the distance between $R_{i,j}$ and the true landmark $T_{i,j}$ is greater than 16 mm in the X-, Y-, or Z-direction, we consider that the RF model cannot easily localize L_i, and L_i is discarded to further trim the candidate landmark set. The remaining 1802 landmarks are kept as the candidate landmark set.

The mean of the coarse positions of landmark L_i is calculated as

$$\bar{R}_i = \frac{1}{80}\sum_{j=1}^{80} R_{i,j}. \tag{1}$$

The maximum Euclidean distance between a landmark in the atlas image and the corresponding landmark that is localized by the RF model in an image in the second training dataset is calculated as

$$\text{MaxDist} = \max_i[\max_j(\|L_i - R_{i,j}\|)], \tag{2}$$

where $i = \{1, 2, \ldots, 1802\}$ and $j = \{1, 2, \ldots, 80\}$. In our experiments, the value of MaxDist is 70 mm, and we use this value to find possible RF model localization errors in the testing phase.

2.5 *Selection of the Most Reliable Landmarks*

We use a RANSAC algorithm to select the most reliable landmarks (denoted as robust set) from the 1802 candidate landmarks (denoted as whole set). The robust set is empty at the beginning. Our algorithm works as follows (Fig. 3):

Step 1: Randomly draw a subset of 18 landmarks (1% of the whole set) from the whole set.

Step 2: Register the atlas image to each of the 80 images in the second training dataset by calculating the TPS transformations from the atlas image to the training images with the subset as control points and project all the landmarks in the whole set from the atlas to each of the 80 volumes with these transformations. Here, we have used a fixed smoothing parameter value of 0.5 for calculating the TPS transformations. The mean and standard deviation of the registration error for each landmark in the whole set are calculated as

$$\bar{\epsilon}_i = \frac{1}{80}\sum_{j=1}^{80} \epsilon_{i,j}, \tag{3}$$

$$\sigma_i = \sqrt{\frac{1}{80}\sum_{j=1}^{80} (\epsilon_{i,j} - \bar{\epsilon}_i)^2}, \tag{4}$$

where the registration error $\epsilon_{i,j}$ of the landmark L_i in image I_j is the Euclidean distance between the true landmark $T_{i,j}$ and the point $P_{i,j}$ obtained by projecting the

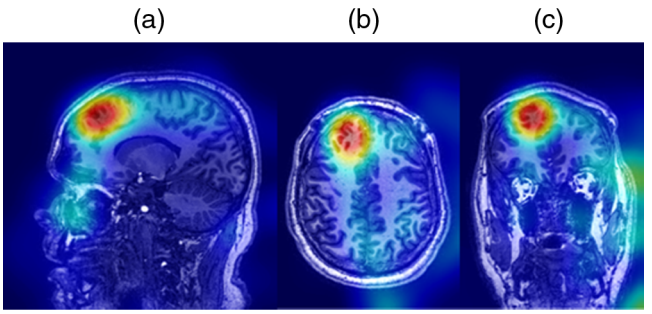

Fig. 2 Probability map shown on top of a training image in the (a) sagittal, (b) axial, and (c) coronal directions. The reddish color indicates higher probability values and the blueish color indicates lower probability values.

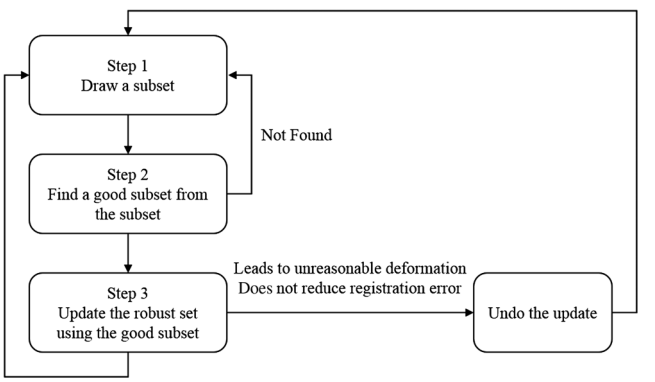

Fig. 3 The flowchart of our RANSAC algorithm.

landmark L_i from the atlas to the image I_i via the TPS transformation. All L_i with $\bar{\epsilon}_i < 8$ mm and $\sigma_i < 4$ mm are considered to be inliers. If the number of inliers is greater than 1261 (70% of the whole set), it suggests that the transformation computed with the landmarks in the current subset is reasonable. Landmarks in the current subset that are also inliers constitute a good subset. If there are less than 1261 inliers, the current subset is discarded and we go back to step 1 to draw a new subset.

Step 3: Update the robust set by adding to it the landmarks in the good subset that are not yet in the robust set.

Step 4: Check the quality of the robust set. To do so, we place a uniform 3-D grid in the atlas image, and the grid coordinates are used as the check set (12,168 points in total). To assess whether or not the TPS transformations calculated with the robust set as control points leads to unreasonable deformations, the determinant of the Jacobian matrix of the TPS transformation at each point in the check set is calculated. This value measures the volume change of a voxel after the TPS transformation, e.g., a value of 0.5 indicates that a unit volume contracts down to half of its original volume, and a value of 2 indicates that a unit volume expands to twice its original volume. Taking into account the fact that the heads in our 201 images are of similar size and position but that large anatomical variation may exist among individuals (especially in the ventricles), we use a very loose decision criterion to determine that a transformation is reasonable, and we consider determinant values between 0.2 and 2.2 to be acceptable. A TPS transformation that leads to valid deformations at more than 99.5% of the check set points is considered to be acceptable. If any TPS transformation that registers the atlas to one of the training images is invalid, then we undo the update of the robust set and go back to step 1; otherwise, we check if the TPS transformations reduce the registration error. To do so, we calculate the sum of the mean and standard registration error as

$$s = \bar{\epsilon} + \sigma, \tag{5}$$

$$\bar{\epsilon} = \frac{1}{1802 \times 80} \sum_{i=1}^{1802} \sum_{j=1}^{80} \epsilon_{i,j}, \tag{6}$$

$$\sigma = \sqrt{\frac{1}{1802 \times 80} \sum_{i=1}^{1802} \sum_{j=1}^{80} (\epsilon_{i,j} - \bar{\epsilon})^2}, \tag{7}$$

if the value of s decreases, then we keep the update; otherwise, we undo the update and go back to step 1.

We repeat steps 1 to 3 until the value of s converges. In our experiments repeated with various seed points for the random number generator used to draw the subset, our algorithm converges with about 500 landmarks in the robust set.

Empirically, we found that too many landmarks affect the results, i.e., too many degrees of freedom may lead to unrealistic transformations when registering the atlas to some unknown new images. We address this by reducing the size of the robust set. To do so, we subdivide the head into a series of 3-D boxes (in the experiment presented here, we subdivide the volume into $4 \times 5 \times 5$ boxes) that cover the whole brain in the atlas. In each of these boxes, we select the landmark that has the smallest sum of the normalized $\bar{\epsilon}_i$ and the normalized σ_i [Eqs. (3) and (4)] if the box contains a landmark. In the end, the trimmed robust set contains 41 points that provide a good coverage of the brain (Fig. 4).

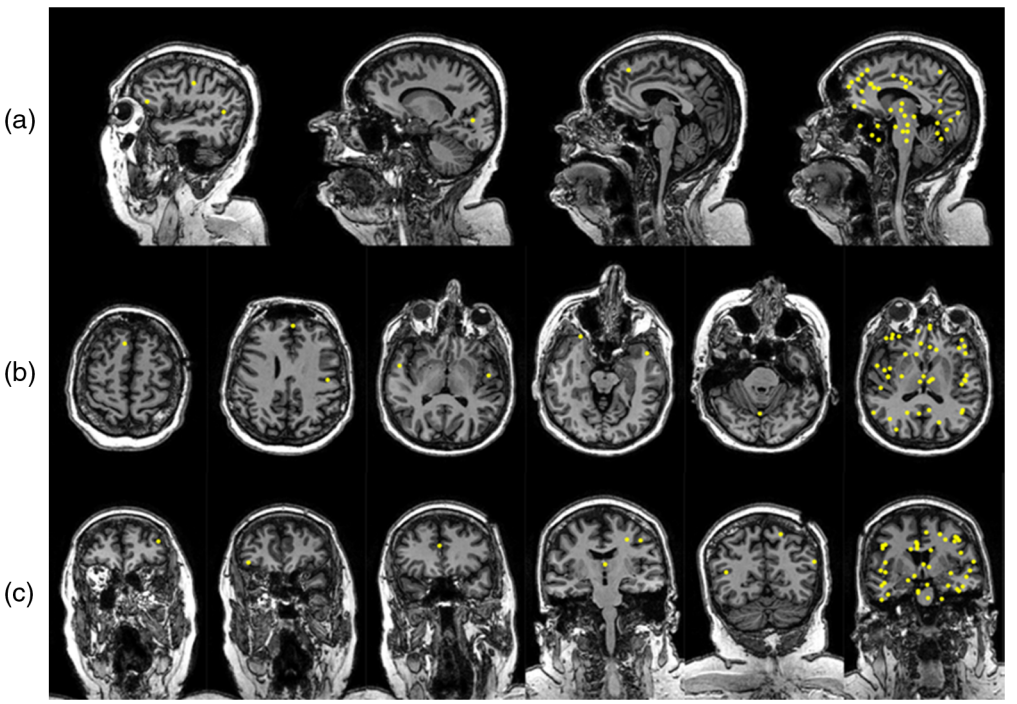

Fig. 4 The robust set shown on selected slices of the atlas image volume in the sagittal (row 1), axial (row 2), and coronal (row 3) directions. The last image of each row is generated by projecting all landmarks on the same slice to show the region of the head covered by the landmarks.

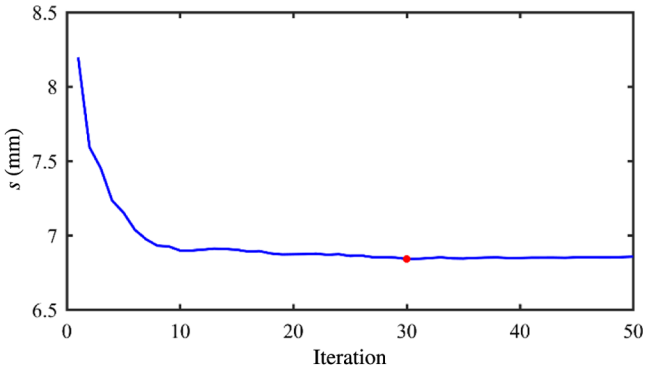

Fig. 5 Registration error s at each iteration.

2.6 *Method Validation*

In the testing phase, first we use the RF models that are created in Sec. 2.3 to localize the robust set in the 20 testing images. As is done in Sec. 2.4, the images in the testing dataset are downsampled by a factor of 4. Given a downsampled testing image and an RF model that is trained to localize a specific landmark, the output of the RF model is a 3-D probability map of the same size as the input image, and the local maxima of the 3-D probability map, which are the centroids of regions above a threshold, are the potential positions of the landmark in the image. Because these points are localized in the downsampled image, we calculate their coarse positions in the full resolution image by upsampling their coordinates. If, for a landmark L_i, multiple points are localized in a testing image, then the point that is the closest to $\bar{R}_i$ [Eq. (1)] is selected. For a testing

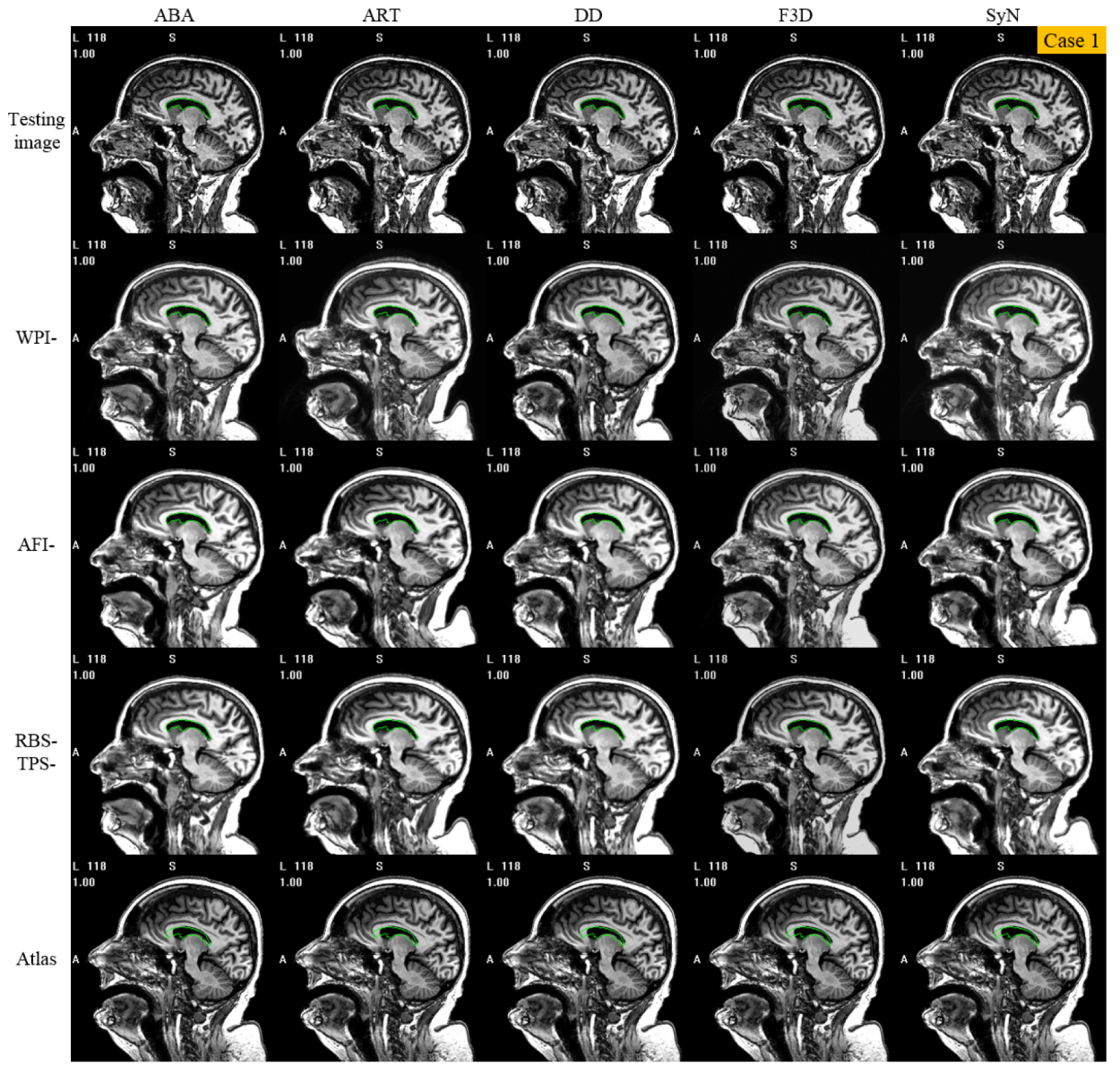

Fig. 6 Sagittal view of example case 1. The testing image (row 1); transformed atlas images using the WPI-, the AFI-, and the RBS-TPS-approaches (rows 2–4, respectively); and the original atlas image (row 5) are shown. Columns 1–5 of rows 2–4 show the transformed atlas images when ABA, ART, DD, F3D, and SyN is used as the deformable registration algorithm. The green contours are drawn on the testing image, and the contours are copied on the transformed atlas and the original atlas images.

image I_j, a small landmark set $\{L_k\}$, in which $\|L_k - R_{k,j}\| >$ MaxDist [Eq. (2)], that may potentially have localization errors are removed from the control points of the TPS registration for this image. For the testing images in our study, the maximum size of $\{L_k\}$ is 4 and the probability of this event is 1%; more frequently, the size of $\{L_k\}$ is 2 or 3.

We use the presented method to compute the initialization transformation for the five deformable registration algorithms. In our experiments, the atlas is first registered to each of the 20 testing images with the TPS-based transformations that use the robust set as control points. Next, these transformed atlas images are registered to each of the testing images with each of the five deformable registration algorithms for further refinement. These registration methods are referred to as RBS-TPS-ABA, RBS-TPS-ART, RBS-TPS-DD, RBS-TPS-F3D, and RBS-TPS-SyN for simplicity. The only nondeterministic factor in our RBS-TPS-method is the random subset, which depends on the initial state of the random number generator that performs the random sampling. To test the sensitivity of our method to the initial state of the random number generator, we run our algorithm with five different initial states, and we use repeated measures ANOVA[16] to assess whether the performance of the RBS-TPS-approach is consistent across the five states.

The validation is performed by comparing our RBS-TPS-approach with four other approaches: (1) applying the deformable registration algorithms without preregistration initialization (referred to as WPI-ABA, WPI-ART, WPI-DD, WPI-F3D, and WPI-SyN), (2) applying the deformable registration algorithms after TPS preregistration initialization using 40 landmarks that are randomly selected from the 1802 candidate landmark set as

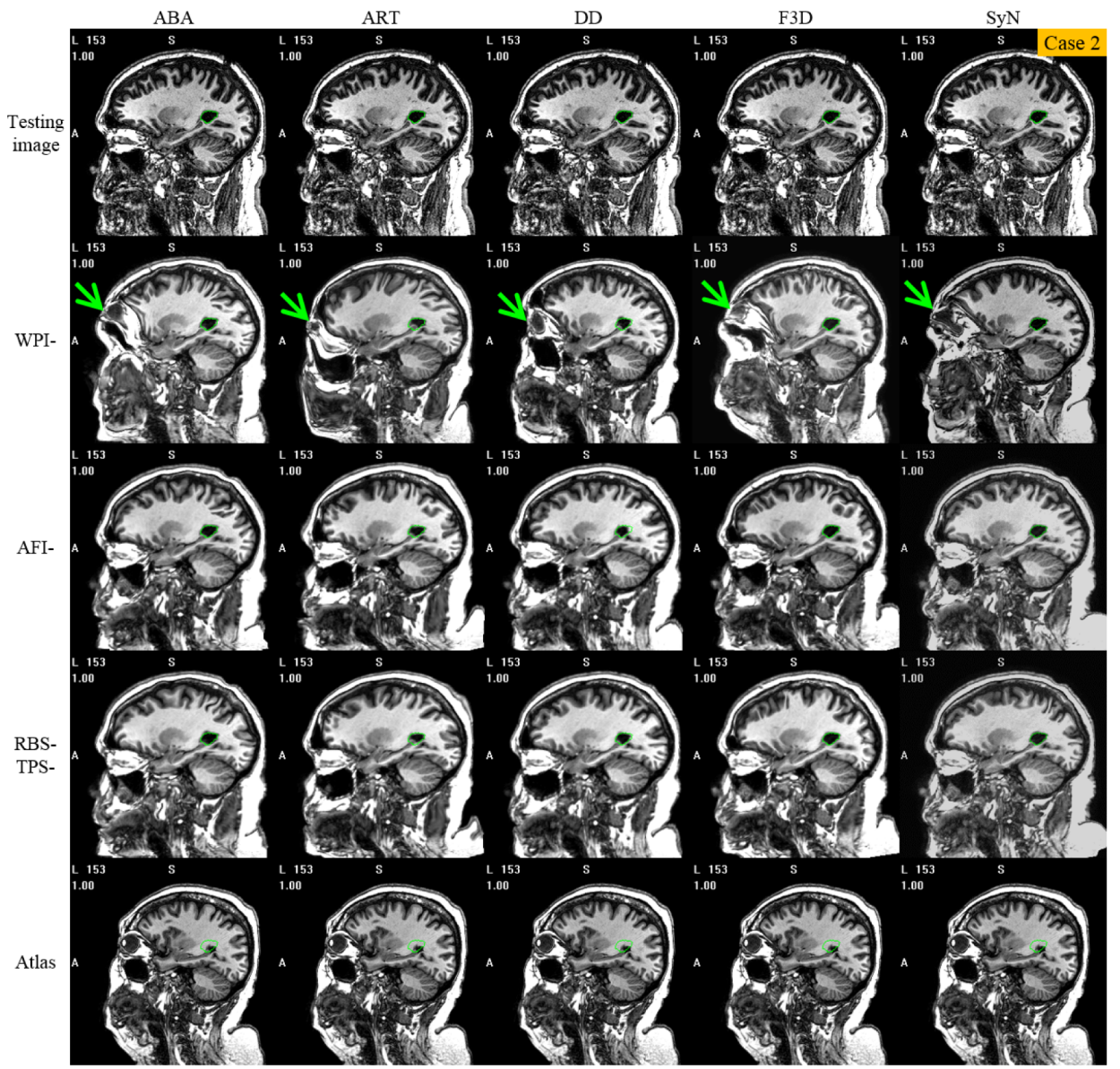

Fig. 7 Sagittal view of example case 2. The testing image (row 1); transformed atlas images using the WPI-, the AFI-, and the RBS-TPS-approaches (rows 2–4, respectively); and the original atlas image (row 5) are shown. Columns 1–5 of rows 2–4 show the transformed atlas images when ABA, ART, DD, F3D, and SyN is used as the deformable registration algorithm. The green contours are drawn on the testing image, and the contours are copied on the transformed atlas and the original atlas images. The green arrows point to regions where the registration failed.

control points (referred to as RND-TPS-ABA, RND-TPS-ART, RND-TPS-DD, RND-TPS-F3D, and RND-TPS-SyN), (3) applying the deformable registration algorithms after TPS preregistration initialization using the 40 landmarks that have the smallest mean registration error [Eq. (3)] as control points (referred to as MINERR-TPS-ABA, MINERR-TPS-ART, MINERR-TPS-DD, MINERR-TPS-F3D, MINERR-TPS-SyN), and (4) applying the deformable registration algorithms after preregistration with an affine transformation computed with a standard intensity-based registration algorithm that uses mutual information as a similarity measure (referred to as AFI-ABA, AFI-ART, AFI-DD, AFI-F3D, and AFI-SyN). The comparison is performed qualitatively and quantitatively. First, we compare the registration results, i.e., the transformed atlas images obtained with the five registration approaches to the testing images. The approach that most often results in a better visual correspondence between the transformed atlas images and the testing images is deemed to be superior to the others. Second, we use the Dice similarity coefficients (DSC)[17] of the gray matter (GM), the white matter (WM), and the cerebrospinal fluid (CSF) between the transformed atlas image and the testing image to quantify the similarity of the two images. To calculate the DSC, the brains of the atlas image and of the 20 testing images are first segmented into GM, WM, and CSF with the FSL automated segmentation tool.[18] The segmented atlas is projected onto each of the segmented volumes in the testing set with the transformations computed in each of the five aforementioned registration approaches. The DSC of tissue class V is calculated as

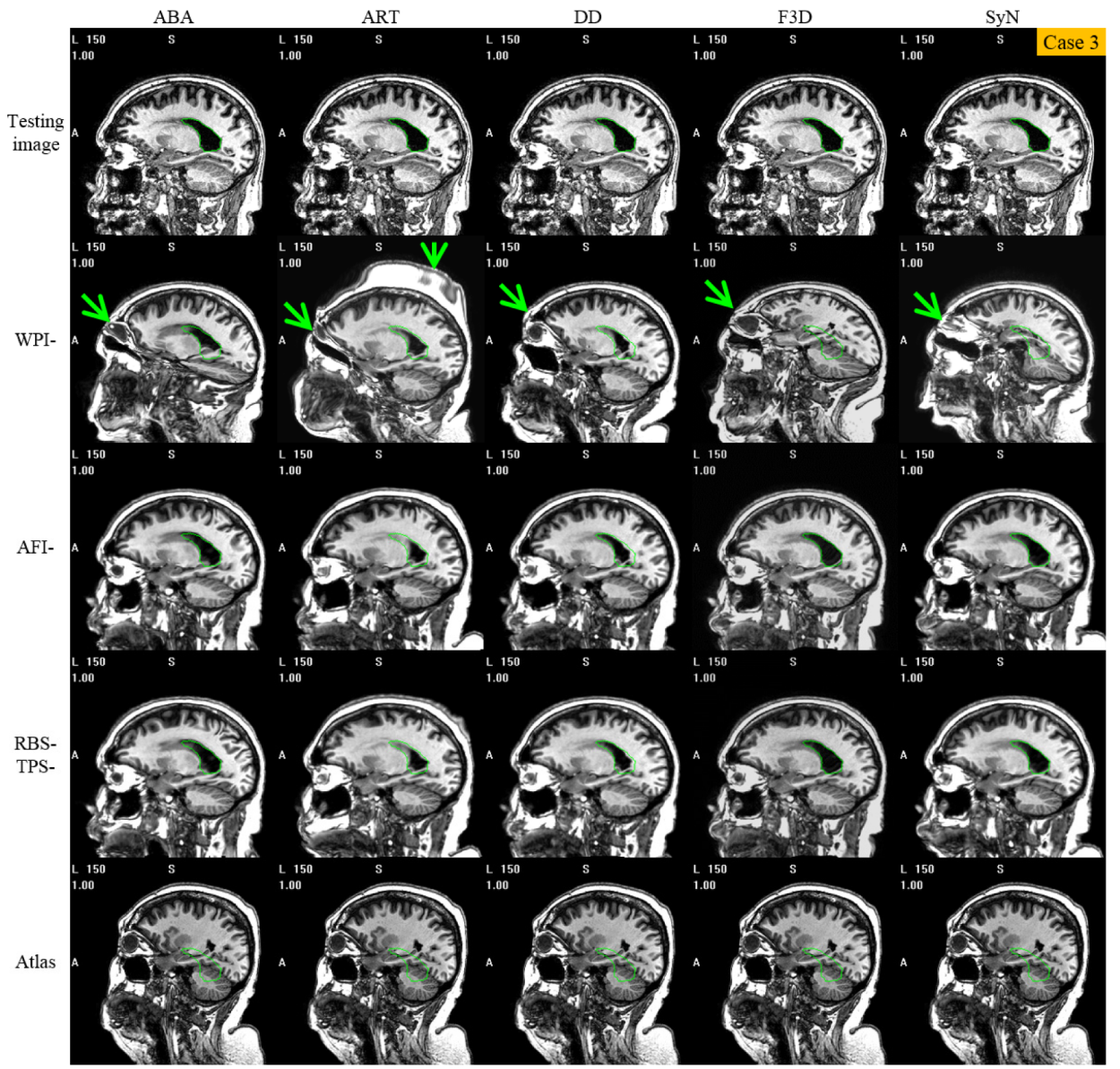

Fig. 8 Sagittal view of example case 3. The testing image (row 1); transformed atlas images using the WPI-, the AFI-, and the RBS-TPS-approaches (rows 2–4, respectively); and the original atlas image (row 5) are shown. Columns 1–5 of rows 2–4 show the transformed atlas images when ABA, ART, DD, F3D, and SyN is used as the deformable registration algorithm. The green contours are drawn on the testing image, and the contours are copied on the transformed atlas and the original atlas images. The green arrows point to regions where the registration failed.

$$\mathrm{DSC}_V = \frac{2 \times |V_{\mathrm{trans}} \cap V_{\mathrm{test}}|}{|V_{\mathrm{trans}}| + |V_{\mathrm{test}}|}, \tag{8}$$

where $V = \{\mathrm{GM}, \mathrm{WM}, \mathrm{CSF}\}$ and V_{trans} and V_{test} denote the voxels with tissue label V in the transformed atlas image and the testing image, respectively. $|V_{\mathrm{trans}}|$ and $|V_{\mathrm{test}}|$ are the numbers of voxels in the two groups, and $|V_{\mathrm{trans}} \cap V_{\mathrm{test}}|$ is the number of overlapping voxels of the two groups. For reference, we also calculate a baseline DSC as

$$\mathrm{DSC}_V = \frac{2 \times |V_{\mathrm{atlas}} \cap V_{\mathrm{test}}|}{|V_{\mathrm{atlas}}| + |V_{\mathrm{test}}|}, \tag{9}$$

where V_{atlas} denotes the voxels with tissue label V in the original atlas image. Finally, paired t-tests[19] are performed to assess whether or not the DSC of the RBS-TPS-approach is statistically significantly different from the DSC of each of the other methods.

3 Results

Figure 5 shows the value of s [Eq. (5)] through iterations of the RANSAC algorithms. Convergence is achieved in about 30 iterations.

Results from the various registration approaches on three cases are shown in Figs. 6–8. These image volumes have been selected from the 20 testing images based on their baseline

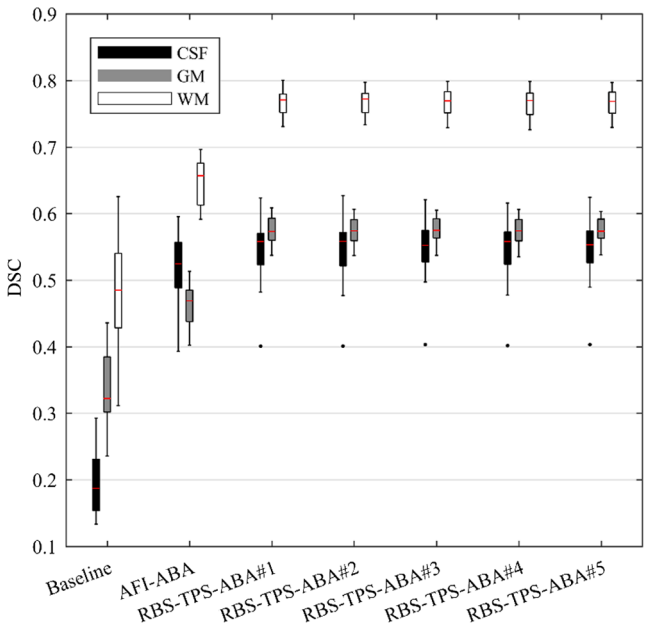

Fig. 9 Box plots of the DSC of CSF, GM and WM regions, for the baseline, AFI-ABA, and RBS-TPS-ABA.

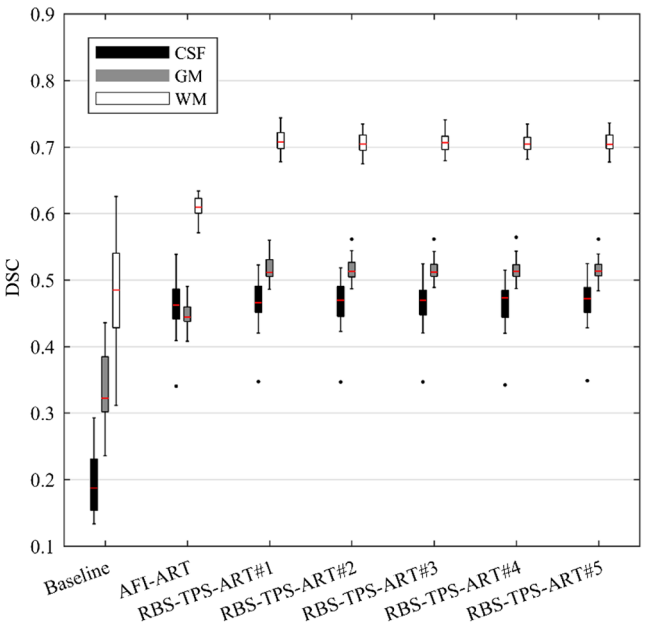

Fig. 10 Box plots of the DSC of CSF, GM and WM regions, for the baseline, AFI-ART, and RBS-TPS-ART.

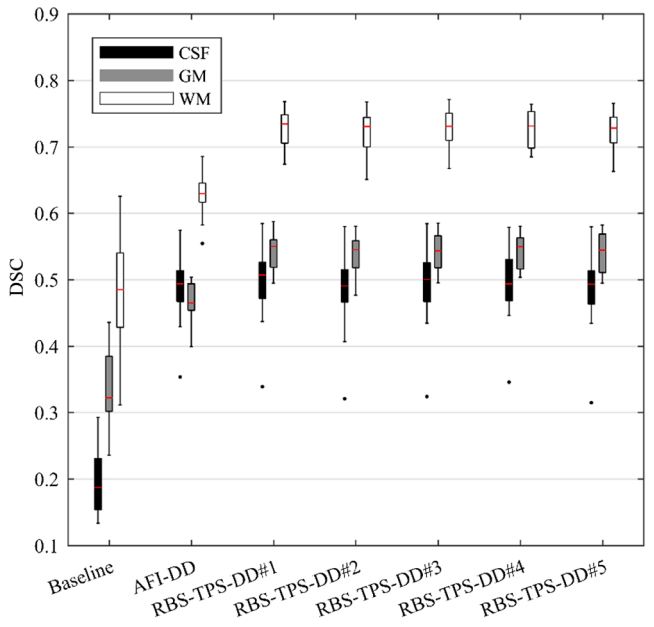

Fig. 11 Box plots of the DSC of CSF, GM and WM regions, for the baseline, AFI-DD, and RBS-TPS-DD.

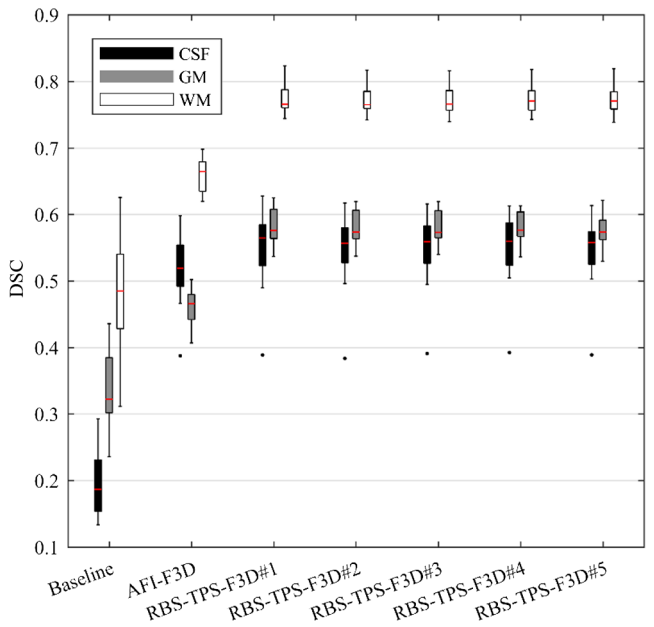

Fig. 12 Box plots of the DSC of CSF, GM and WM regions, for the baseline, AFI-F3D, and RBS-TPS-F3D.

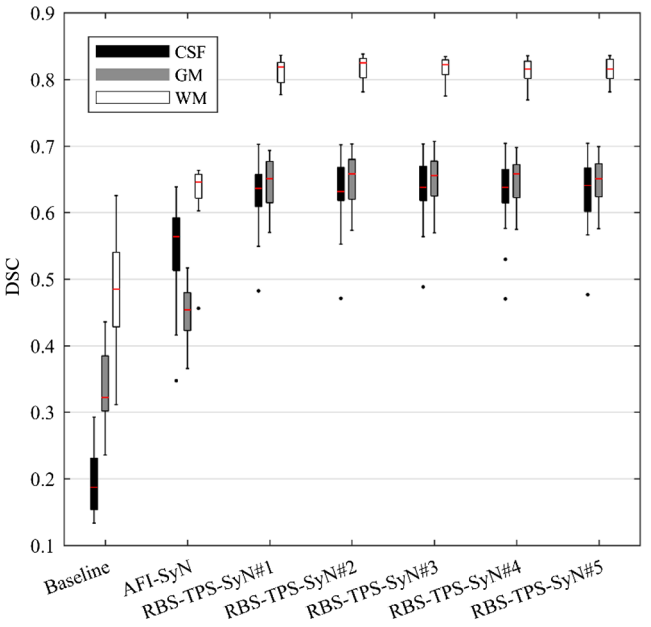

Fig. 13 Box plots of the DSC of CSF, GM and WM regions, for the baseline, AFI-SyN, and RBS-TPS-SyN.

DSC values. Case 1 has the highest baseline DSC among the 20 cases, case 2 has the baseline DSC value that is closest to the mean baseline DSC value of the 20 cases, and case 3 has the lowest baseline DSC value among the 20 cases. The accuracy of the registration obtained with WPI-ABA, WPI-ART, WPI-DD, WPI-F3D, and WPI-SyN, i.e., when the deformable registration method is applied without initialization, is visually assessed, and we observe a failure rate of at least 20% for each algorithm. This confirms that deformable registration methods require a good preregistration initialization. The failure of the WPI-approaches is apparent in the frontal lobe regions (arrows on Figs. 7 and 8, row 2) of cases 2 and 3. On the same image volumes, both the AFI-approach and the RBS-TPS-approach lead to good results (Figs. 7 and 8, rows 3 and 4). The ventricular region in the testing image is delineated (Figs. 6–8, row 1, green contours), and the contours are copied on all the other images. This visually shows that the ventricles are accurately registered and that our RBS-TPS-approach produces results that are comparable to the standard AFI-approach for these structures.

Table 3 *P*-values of the repeated measures ANOVA of the RBS-TPS-approaches with five different initial states.

Approach	WM	GM	CSF
RBS-TPS-ABA	0.694	0.725	0.183
RBS-TPS-ART	**<0.01**	0.094	0.274
RBS-TPS-DD	0.127	0.511	0.005
RBS-TPS-F3D	0.055	0.150	0.518
RBS-TPS-SyN	0.095	0.065	0.063

Note: Bold indicates cases that are significantly different (p-value less than 0.01).

Figures 9–13 show box plots of the DSC for CSF (black), GM (gray), WM (white), for the baseline, the AFI-approach, and the RBS-TPS-approach with five different initial states. The same trend is observed for all five algorithms, i.e., the DSC values of the RBS-TPS-approaches are higher than those of the AFI-approaches. Paired *t*-tests (Table 2) show that for ABA, F3D, and SyN, the RBS-TPS-approach results in statistically significant higher DSC for WM, GM, and CSF than the AFI-approach ($p < 0.01$ for WM, GM, and CSF); for ART and DD, the RBS-TPS-approach results in statistically significant higher DSC for WM and GM than the AFI-approach ($p < 0.01$ for WM and GM); there is a substantial but not statistically significant difference between the DSC for CSF obtained with the two approaches ($p > 0.01$). Repeated measures ANOVA (Table 3) shows that statistically significant differences do not exist in the DSC of the five RBS-TPS-trials that are conducted with five different initial states, except for ART ($p < 0.01$ for WM). This suggests that our RBS-TPS-approach is robust against the initial state of the random number generator for four of the five registration methods that we test. For ART, despite being statistically significant for the WM, the difference is small.

Experiments show that the RND-TPS-approach and the MINERR-TPS-approach are not feasible. The TPS transformation fails or results in unreasonable deformation for some testing images when randomly selected landmarks are used as control points. The TPS transformation fails on 100% of our testing images when the landmarks that have the smallest mean registration error are used as control points. This is not unexpected because we have observed that these landmarks are typically

Table 2 *P*-values of the paired *t*-test between DSC for WM, GM, and CSF for the RBS-TPS-approaches and the AFI-approaches.

	RBS-TPS-M#1			RBS-TPS-M#2			RBS-TPS-M#3			RBS-TPS-M#4			RBS-TPS-M#5		
M	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
ABA	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
ART	<0.01	<0.01	**0.33**	<0.01	<0.01	**0.62**	<0.01	<0.01	**0.66**	<0.01	<0.01	**0.80**	<0.01	<0.01	**0.38**
DD	<0.01	<0.01	**0.47**	<0.01	<0.01	**0.66**	<0.01	<0.01	**0.86**	<0.01	<0.01	**0.65**	<0.01	<0.01	**0.76**
F3D	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
SyN	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

M, registration method.
Note: Bold indicates cases that are not significantly different.

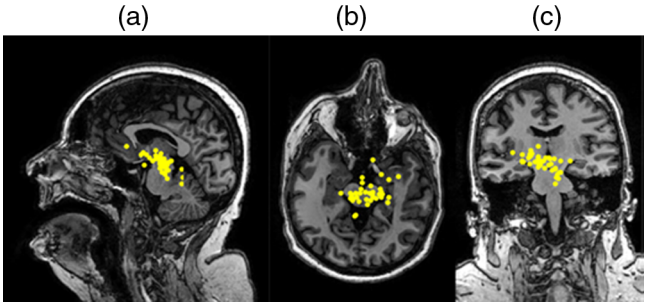

Fig. 14 Landmarks that have the smallest mean registration error are shown on the same slice of the atlas image volume in the (a) sagittal, (b) axial, and (c) coronal directions. All these landmarks are projected on the same slice to show their coverage range.

located near the midbrain; thus, they cannot provide good coverage of the brain (Fig. 14).

4 Conclusions

We present an approach for the selection of reliable landmarks for deformable registration initialization that uses an RF approach followed by a RANSAC step. The method that we proposed is fully automatic and generic. It could be applied to other registration problems. We evaluated our approach using it to initialize five well-established deformable registration algorithms, and our results show that the same trend is observed for all five algorithms, i.e., the final registration results are statistically better with our approach than with a standard approach that relies on the estimation of an affine transformation computed with an intensity-based approach.

Because this approach operates on principles that are different from most nonrigid registration methods in routine use, it could also be used as an error detection mechanism. In this context, it could be run in parallel with existing processing pipelines, and differences observed between methods in either deformation fields or landmark position could trigger alerts; this will be explored in future studies.

Disclosures

The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest.

Acknowledgments

This research was supported in part by the NIH (Grant No. R01-NS095291) and used the resources of the Center for Research and Education at Vanderbilt University, Nashville, Tennessee. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the NIH.

References

1. J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.* **2**(1), 1–36 (1998).
2. D. Han et al., "Robust anatomical landmark detection with application to MR brain image registration," *Comput. Med. Imaging Graph.* **46**, 277–290 (2015).
3. L. Breiman, "Random forests," *Mach. Learn.* **45**(1), 5–32 (2001).
4. K. Rohr et al., "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Trans. Med. Imaging* **20**(6), 526–534 (2001).
5. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), 381–395 (1981).
6. G. K. Rohde et al., "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Trans. Med. Imaging* **22**(11), 1470–1479 (2003).
7. B. A. Ardekani et al., "Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans," *J. Neurosci. Methods* **142**(1), 67–76 (2005).
8. T. Vercauteren et al., "Diffeomorphic demons: efficient non-parametric image registration," *NeuroImage* **45**(1 Suppl.), S61–S72 (2009).
9. M. Modat et al., "Fast free-form deformation using graphics processing units," *Comput. Methods Programs Biomed.* **98**(3), 278–284 (2010).
10. B. B. Avants et al., "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.* **12**(1), 26–41 (2008).
11. P. F. D'Haese et al., "CranialVault and its CRAVE tools: a clinical computer assistance system for deep brain stimulation (DBS) therapy," *Med. Image Anal.* **16**(3), 744–753 (2012).
12. Y. Liu, R. F. D'Haese, and B. M. Dawant, "Effects of deformable registration algorithms on the creation of statistical maps for preoperative targeting in deep brain stimulation procedures," *Proc. SPIE* **9036**, 90362B (2014).
13. S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapp.* **17**(3), 143–155 (2002).
14. B. Glocker et al., "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in *Proc. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. **III**, pp. 590–598 (2012).
15. O. Pauly and B. Glocker, "Fast multiple organs detection and localization in whole-body MRI Dixon sequences," in *Proc. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. **III**, pp. 239–247 (2011).
16. S. D. Charles, *Statistical Methods for the Analysis of Repeated Measurements*, pp. 103–123, Springer-Verlag, New York (2012).
17. T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons," *Biol. Skr.* **5**, 1–34 (1948).
18. Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imaging.* **20**(1), 45–57 (2001).
19. B. L. Welch, "The generalization of 'student's' problem when several different population variances are involved," *Biometrika* **34**(1–2), 28–35 (1947).

Jianing Wang received her bachelor's degree in biomedical engineering from the Southern Medical University, China, in 2010 and her master's degree in biomedical imaging and signals from Illinois Institute of Technology in 2012. She is a PhD student at Vanderbilt University. She was a research assistant in the Department of Radiology at the University of Chicago from 2012 to 2015. Her current research interests include medical image processing, computer-aided detection, and machine learning.

Yuan Liu received her BEng degree in computer science from Northwestern Polytechnical University, Xi'an, China, in 2010, and her PhD at the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee, USA, in 2016. Currently, she is a software engineer at Google Inc. Her research interests include image processing, computer vision, machine learning, and big data analytics.

Jack H. Noble received his BE, MS, and PhD degrees in electrical engineering from Vanderbilt University, Nashville, Tennessee, in 2007, 2008 and 2011, respectively. Currently, he is an assistant professor in the Department of Electrical Engineering and Computer Science, Vanderbilt University. His primary research interests include medical image processing, image segmentation, registration, statistical modeling, and image-guided surgery techniques.

Benoit M. Dawant received his MSEE degree from the University of Louvain, Leuven, Belgium, in 1983 and his PhD from the University of Houston, Houston, Texas, USA, in 1988. He is currently a Cornelius Vanderbilt professor in the Department of Electrical Engineering and Computer Science at Vanderbilt University. His primary research areas include medical image processing, segmentation, and registration. He currently focuses on applying these methods to the development of innovative techniques for guiding surgical and interventional procedures.