Int J CARS
DOI 10.1007/s11548-012-0675-8

ORIGINAL ARTICLE

Automated segmentation of basal ganglia and deep brain structures in MRI of Parkinson's disease

Claire Haegelen · Pierrick Coupé · Vladimir Fonov · Nicolas Guizard · Pierre Jannin · Xavier Morandi · D. Louis Collins

Received: 14 September 2011 / Accepted: 17 February 2012

Abstract
Purpose Template-based segmentation techniques have been developed to facilitate the accurate targeting of deep brain structures in patients with movement disorders. Three template-based brain MRI segmentation techniques were compared to determine the best strategy for segmenting the deep brain structures of patients with Parkinson's disease.
Methods T1-weighted and T2-weighted magnetic resonance (MR) image templates were created by averaging MR images of 57 patients with Parkinson's disease. Twenty-four deep brain structures were manually segmented on the templates. To validate the template-based segmentation, 14 of the 24 deep brain structures from the templates were manually segmented on 10 MR scans of Parkinson's patients as a gold standard. We compared the manual segmentations with three methods of automated segmentation: two registration-based approaches, automatic nonlinear image matching and anatomical labeling (ANIMAL) and symmetric image normalization (SyN), and one patch-label fusion technique. The automated labels were then compared with the manual labels using a Dice-kappa metric and center of gravity. A Friedman test was used to compare the Dice-kappa values and paired *t* tests for the center of gravity.
Results The Friedman test showed a significant difference between the three methods for both thalami ($p < 0.05$) and not for the subthalamic nuclei. Registration with ANIMAL was better than with SyN for the left thalamus and was better than the patch-based method for the right thalamus.
Conclusion Although template-based approaches are the most used techniques to segment basal ganglia by warping onto MR images, we found that the patch-based method provided similar results and was less time-consuming. Patch-based method may be preferable for the subthalamic nucleus segmentation in patients with Parkinson's disease.

Keywords Basal ganglia · MRI template · Patch-based method · Parkinson's disease · Segmentation

C. Haegelen · P. Coupé · V. Fonov · N. Guizard · D. L. Collins
McConnell Brain Imaging Centre, Montreal Neurological Institute, 3801 University Street, Montreal, QC H3A 2B4, Canada

C. Haegelen · P. Jannin · X. Morandi
Faculty of Medicine, U746, INSERM, Rennes Cedex, France

C. Haegelen · P. Jannin · X. Morandi
VisAGeS Unit/Project, INRIA, Rennes Cedex, France

C. Haegelen · P. Jannin · X. Morandi
UMR 06074, IRISA, CNRS, University of Rennes 1, CS 34317, 35043 Rennes Cedex, France

C. Haegelen (✉)
Service de Neurochirurgie, Hôpital Pontchaillou, rue Henri Le Guilloux, 35033 Rennes Cedex 9, France
e-mail: Claire.HAEGELEN@chu-rennes.fr

P. Coupé
LaBRI CNRS, UMR 5800, University of Bordeaux, 351 cours de la Liberation, 33045 Talence Cedex, France

Introduction

Many movement disorders are related first to Parkinson's disease, with symptoms that include essential tremor, rigidity, and akinesia, and less frequently to dystonia that is a syndrome of sustained muscle contractions producing writhing movements and abnormal postures, to essential tremor, or to Tourette syndrome. When pharmaceutical treatments lose effectiveness, such disorders may require functional neurosurgery. In such surgeries, deep brain electrodes are implanted to inhibit the activity of target structures such as the subthalamic nucleus (STN), medial globus pallidus (GPm),

Published online: 18 March 2012

and caudal part of the thalamic ventral lateral nucleus (VLc) [1–3]. Following surgery, motor effects and neuropsychological or psychiatric secondary side effects of the stimulation may occur depending on the location and trajectory of the electrodes [3–5]. The patient's surgical outcome is related to the accuracy of nucleus targeting during surgery, and the accuracy of electrode implantation depends on different steps such as image preprocessing, intraoperative clinical testing of patients, and intraoperative electrophysiological recordings.

Brain atlases and atlas-based segmentation techniques have been developed to facilitate the accurate targeting of deep brain structures in patients with movement disorders [6–11]. Recently, several authors have demonstrated improved targeting accuracy by using multimodal databases composed of anatomical and functional data [12–14]. In these studies, basal ganglia segmentations, the coordinates of previously implanted electrodes, and functional data (mainly intracerebral electrophysiological recordings) were aligned in a common space and overlaid on a magnetic resonance (MR) image template. Although the use of multimodal images has been shown to be an effective tool in surgical planning, very few studies have correlated clinical postoperative data with anatomical data [13].

Developing a tool to improve targeting accuracy and the aggregation of anatomical and clinical data requires several steps, including the construction of a template to define a common stereotactic space, accurate and automated segmentation of target structures on the template, registration of the patient's multimodal image data, and alignment of the patient's image data into a common space for comparison with electrode placement. In this paper, we focus on the segmentation step by investigating the accuracy of three methods of segmenting the basal ganglia and deep brain structures of patients with Parkinson's disease. The goal of this work was to compare these methods and determine which method was enough accurate and clinically applicable to obtain segmentation of target structures in further patients with Parkinson's disease waiting for deep brain stimulation.

First, we used two segmentation methods that are nonlinear registration-based approaches involving a population-specific, anatomically labeled template. This kind of approach is widely used and has demonstrated good performance in the segmentation of a large variety of structures [15]. In our study, we constructed a population-specific template [16], which has been shown to be one of the most effective methods of achieving accurate segmentation of subjects with pathology. To achieve automated segmentation, the basal ganglia and other deep brain structures useful in the planning of deep brain stimulation were first manually delineated on the template, which was then nonlinearly warped onto the patient's MR image. The segmentation of the target structures on the MR image could thus be obtained by mapping the template labels through the estimated transformation onto the patient's MR image. We compared two nonlinear algorithms to achieve this task: the well-known automatic nonlinear image matching and anatomical labeling (ANIMAL) method [15] and the more recent symmetric image normalization (SyN) technique [17]. The third segmentation method evaluated here was a recently published, patch-based, label fusion technique that has demonstrated very good performance for hippocampus and ventricle segmentation [18]. We compared the results of the three segmentation methods in terms of anatomical similarity with the template segmentation and discussed the interest of these methods in clinical practice for deep brain stimulation.

Materials and methods

Patients

The template was constructed using T1-weighted (T1w) and T2-weighted (T2w) MR images from 57 patients with Parkinson's disease. The patients (26 women and 31 men; mean age [±standard deviation (SD)] 58.9 ± 8 years; mean disease duration 12.2 ± 5 years) underwent electrode implantation in the STN, GPm, or VLc according to their disease (Table 1).

Before imaging, a Leksell stereotactic frame (Elekta Instruments AB, Stockholm, Sweden) was rigidly fixed on the patient's head. All MR images were acquired on a Philips Achieva 3T scanner (Philips Medical Systems, Best, The Netherlands) using a transmit-and-receive head coil. After intravenous injection of gadolinium (0.2 ml/kg), the T1w data were acquired using a three-dimensional (3D) axial

Table 1 Demographic and clinical data for 57 patients with Parkinson's disease

Target	Side	Number of patients	Mean age (years)	Sex	Mean disease duration (years)
VLc	1 bilateral 4 left/2 right	7	65.1	3 F, 4 M	12.4
GPm	24 bilateral	24	60.2	12 F, 12 M	12.6
STN	26 bilateral	26	56.1	11 F, 15 M	11.7
Total		57	58.9	26 F, 31 M	12.2

VLc caudal part of the thalamic ventral lateral nucleus; *GPm* medial globus pallidus; *STN* subthalamic nucleus; *M* male; *F* female

fast field echo sequence (TR = 9.8 ms, TE = 4.6 ms, flip angle = 8°, 256 mm field of view, 219 continuous slices, 1 mm thickness, 256 × 256 matrix upsampled to voxel size of 0.5 × 0.5 × 1 mm). For 42 of the 57 patients, a T2w coronal turbo spin echo sequence was acquired (TR = 3, 035 ms, TE = 80 ms, flip angle = 90°, 36 continuous slices, 1 mm thickness, 256 × 256 matrix upsampled to voxel size of 0.5 × 0.5 × 1 mm). The T2w image volume was centered on the thalamus. The study was approved by the local research ethics committee, and informed consent was obtained from all participants.

Construction of the Parkinson's templates

We used the procedure detailed by Fonov et al. [16] to construct multispectral, multisubject, unbiased nonlinear average Parkinson's disease templates. The 57 T1w images were used to create the T1w Parkinson's template (PD_T1_57), and the 42 T2w images from the same subjects were used to create the T2w Parkinson's template (PD_T2_42), using the steps described below. All images were converted to the Medical Image NetCDF (MINC) format [19], the format used at our institute for image processing.

The following preprocessing steps were applied to all MR scans prior to constructing the template: (1) the program N3 was used to correct for image intensity nonuniformity [20]; (2) inter-subject linear image intensity normalization was achieved using linear histogram scaling to the International Consortium Brain Mapping (ICBM152) template [16] following the technique of Nyul et al. [21]; (3) intrasubject inter-modality volume registration of T2w and T1w MR images was achieved with a rigid-body transformation (i.e., only 3 translations and 3 rotations) and a mutual-information cost function [22] implemented in the *mritoself* tool from the MINC mni_autoreg software package [23]; (4) inter-subject spatial alignment to the Talairach-like ICBM152 stereotactic space was achieved with a linear (nine parameter) transformation estimated with *mritotal*; and (5) a patient-specific brain mask was created by applying brain extraction tool (BET) from the Oxford centre of Functional MRI of the brain functional software library (FSL) to the T1w data [24]. Then, the T1w and T2w data for all subjects were intensity-normalized and spatially aligned within the ICBM152 stereotactic space.

After preprocessing, the average high-signal-to-noise T1w and T2w templates were created using an iterative unbiased method [16] (see Fig. 1). At each iteration, all T1w datasets are nonlinearly registered to an intermediate target and then averaged together on a voxel-by-voxel basis. The inverse of the average nonlinear transformation is applied to the intermediate template to remove any potential spatial bias. This procedure is repeated multiple times in a hierarchical fashion, gradually increasing the resolution of the nonlinear transformations, with number of iterations estimated empirically to achieve good results from a similar dataset. The result is an unbiased, high-contrast, highly detailed T1w average MR image template. Nonlinear registrations were obtained with the ANIMAL procedure [15], chosen because of our extensive experience with this technique. Following the unbiased nonlinear registration, the transformations estimated on the T1w data were applied to the patient's corresponding T2w data to create the unbiased T2w average template. The templates (PD_T1_57 and PD_T2_42) enable the visualization of deep brain structures such as amygdala or hippocampus and basal ganglia structures such as subthalamic nucleus or thalamus. These structures have either T1w contrast or T2w contrast, or are visible in both modalities, and can be manually segmented on each slice of the PD_T1_57 and/or PD_T2_42 templates.

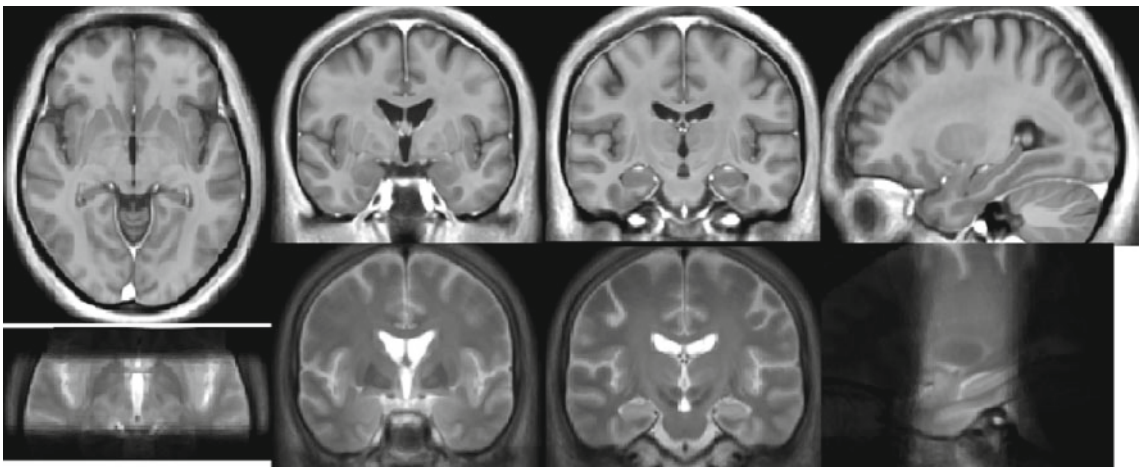

Fig. 1 Parkinson's templates obtained from 57 T1w and 42 T2w average MR images of patients with Parkinson's disease. From *left* to *right*: Axial, coronal, coronal and sagittal slices of the T1w (*top*) and T2w (*bottom*) templates

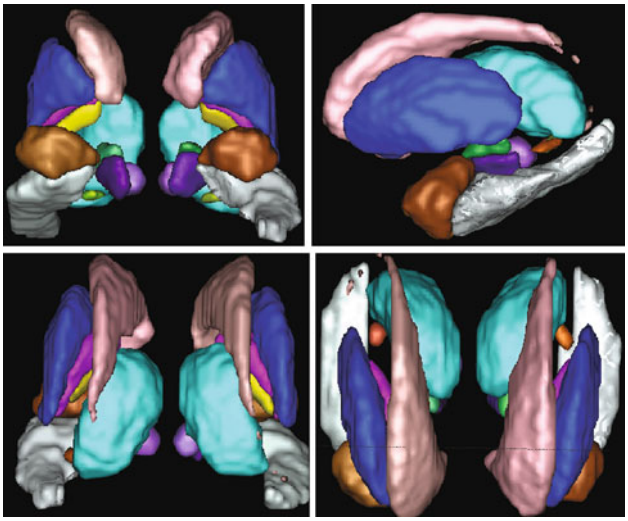

Fig. 2 Anterior (*top left*), lateral left (*top right*), posterior (*bottom left*), and superior (*bottom right*) 3D views of the manual segmentation of the Parkinson's templates. The painted structures included bilateral amygdala (*brown*), hippocampus (*white*), putamen (*dark blue*), GPm (*yellow*), lateral globus pallidus (*dark pink*), caudate nucleus (*light pink*), thalami (*light blue*), substantia nigra (*dark violet*), STN (*dark green*), red nucleus (*light violet*), and medial geniculate body (*light green*)

Manual segmentation

Manual segmentation of the template was achieved using the MNI Display software (Montreal Neurological Institute, Canada, www.bic.mni.mcgill.ca/ServicesSoftware) by the first author, Claire Haegelen. Segmentation consisted of manual delineations of 24 structures bilaterally on each slice of the PD_T1_57 and PD_T2_42 templates. The Display software allows the user to paint the structure in any chosen slice (e.g., transverse) with simultaneous update in the other two views (e.g., coronal and sagittal), enabling consistent segmentation in 3D while viewing either template. The 24 structures were identified with the help of anatomical atlases [9,10,25]. The structures were defined bilaterally and included the amygdala, hippocampus, caudate nucleus, putamen, GPm and lateral globus pallidus, thalamus, medial, and lateral geniculate bodies, red nucleus, substantia nigra, and STN (see Fig. 2). All the structures were visible on the PD_T2_42 template. While a rapid 36-slice T2w image centered on the thalamus was acquired on the morning of the PD surgery, it did not permit to identify for segmentation the entire caudate nucleus, amygdala, and hippocampus. For this reason, these structures were segmented on the PD_T1_57 template where they were visible. Display allows both T1w and T2w images to be visible simultaneously to ensure label consistency. It took 120 h to manually segment the structures on the templates.

To validate the automated template-based segmentation procedure, a gold standard labeling of patient data was required. Fourteen bilateral structures were manually painted on the T1w and T2w images of 10 patients with Parkinson's disease. The images used were a subset of those used for the Parkinson's templates after N3 nonuniformity correction [20]. Using the Display program, the structures labeled included the amygdala, hippocampus, putamen, thalamus, red nucleus, substantia nigra, and STN (see Fig. 3). As for the template, the amygdala and hippocampus were segmented on the patients' T1w image, and the remaining structures on the T2w image. We chose these nine structures for their different sizes: large (e.g., thalamus and hippocampus), intermediate (e.g., amygdala and putamen), and small (e.g., red nucleus, substantia nigra, and STN). The five remaining structures (caudate nucleus, GPm, and lateral globus pallidus, medial and lateral geniculate bodies) were not identified on the patient scans.

To measure the quality of the manual segmentation used as a gold standard, two experts (Claire Haegelen and Louis Collins) segmented the putamen, subthalamic nucleus, and substantia nigra, on both the left and right sides, on the MRI of one patient. These segmentations were used to estimate the intra- and inter-rater variability in structure segmentation.

Registration with ANIMAL and SyN

As described earlier, the registration-based segmentation procedure maps the labels from the PD_T1_57 template onto the MR image using the nonlinear registration transformation that aligns the template and patient MR image data. The template's labels, customized to the anatomy of each patient, can then be compared with the gold standard labels described above. Two methods of inter-subject nonlinear registration were evaluated: ANIMAL [15] and SyN [17]. So, as not to bias the comparison between the two methods, only the T1w modality was used to drive the nonlinear registration. Furthermore, the PD_T1_57 template was more appropriate for the whole-brain registration since the PD_T2_42 only covered a part of the brain.

Linear and nonlinear transformations were calculated between the PD_57_T1 template and each patient's T1w image with both ANIMAL and SyN, resulting in two spatial warping transformations. Then, the template's labels were mapped through each of the two transformations onto the patient's T1w image to achieve two automatic structure segmentations: one for ANIMAL and one for SyN (see Fig. 4).

Patch-based segmentation method

In addition to the registration-based methods, we also evaluated a novel patch-based method using anatomical information derived from manual segmentations by experts, as described in detail by Coupe et al. [18]. This method required only linear registration into the MNI space of the involved

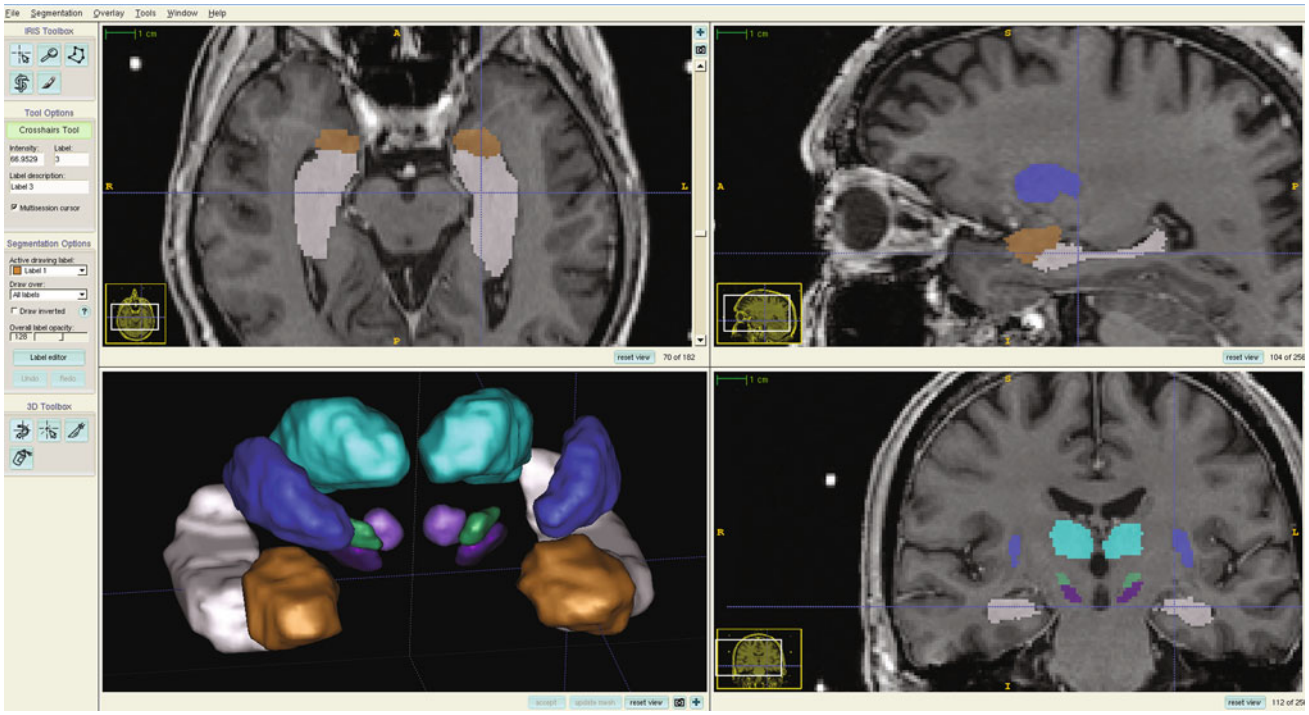

Fig. 3 Screenshot of a patient's manual painting using ITKsnap. Axial (*top left*), sagittal (*top right*), and coronal (*bottom right*) slices of MR image and 3D (*bottom left*) view of the following bilateral painted structures: amygdala (*brown*), hippocampus (*white*), putamen (*dark blue*), thalami (*light blue*), substantia nigra (*dark violet*), STN (*dark green*), and red nucleus (*light violet*)

images (i.e., subject under study and the training templates) as well as a denoising step with the 3D blockwise nonlocal means filter [26,27]. In brief, for each voxel in the MR image that needs to be segmented, the neighborhood of the voxel is compared with similar neighborhoods in previously labeled MR image datasets that serve as a library of templates. A weight, proportional to the neighborhood similarity between the image and the templates, is computed and used as a label voting mechanism to achieve segmentation. In this manner, anatomically similar neighborhoods from one or more templates in the library can contribute to the labeling of the patient's MR image. Here, the gold standard labels are used to define the template library, and experiments are completed in a leave-one-out fashion (i.e., the patient being tested is removed from the template library, and the remaining patients form the library).

For each patient, the patch-based method was used to obtain the same labels as those identified manually for the gold standard. We first processed the method on the T1w images of the 10 patients, but the substantia nigra, red nuclei, and STN were only visualized on the T2w sequences. For these six labels, the patch-based method was applied to the T2w images using the T2w data from the gold standard template library. Finally, the eight labels obtained from the T1w patch-based procedure and those obtained from the T2w patch-based procedure were combined to complete the patch-based segmentation for each patient. Figure 4 presents a view of the different labels obtained with the patch-based method.

Quality of fit metrics

Finally, the automatic labels obtained with either ANIMAL, SyN, or the patch-based method were compared with the manual labels of the gold standard using two metrics: the Dice's kappa similarity coefficient (kappa) [28] because this metric is well established in the literature for comparison between two segmentations and a center-of-gravity metric because this will yield a more intuitive estimate of the potential targeting error. We did not estimate a Hausdorf distance or edge-based/border-based distances because these can be too variable, as they are driven by extreme outlier values. The kappa was measured as follows:

$$K = 2 * (V(M \cap A))/(V(M) + V(A))$$

where M is the set of manually labeled voxels, A is the set of automatically labeled voxels, $\cap$ is the set intersection operator, and $V(\bullet)$ is the volume operator. K takes on a value between 0 and 1.0, with 1.0 indicating perfect agreement. Before the kappa metric is used to evaluate the automated methods, it was first used to estimate the intra- and inter-rater variability of manual structure segmentation to establish a baseline for comparison.

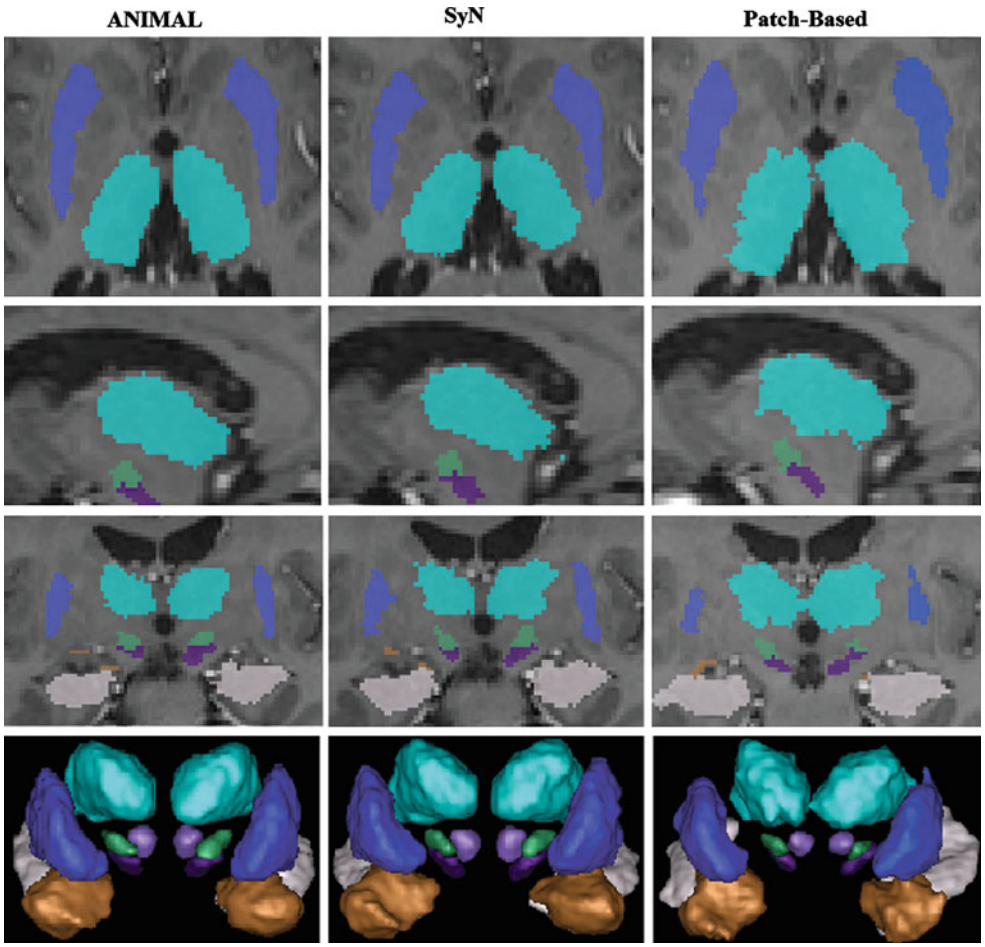

Fig. 4 Generation of labels by ANIMAL registration, SyN registration, and patch-based method. All labels are shown on axial (*top row*), parasagittal (*second row*), and coronal (*third row*) slices of a T1w map and on a 3D reconstruction (*bottom row*). Labeled structures: bilateral amygdala (*brown*), hippocampus (*white*), putamen (*dark blue*), thalami (*light blue*), substantia nigra (*dark violet*), STN (*dark green*), and red nuclei (*light violet*)

The center of gravity was estimated for each segmented structure obtained with the 3 methods, the ANIMAL and SyN registrations and the patch-based method. The Euclidean co-ordinates of the center of gravity were used to estimate a distance between the automatic segmentations and their respective gold standard. A paired-t test was used to compare the distances obtained with ANIMAL and SyN and with the patch-based method. A p-value of less than 0.05 was deemed significant. We also computed the center-of-gravity values between the repeated manual segmentations of the two experts, to estimate inter-rater variability.

Data analysis

The resulting kappa values were not Gaussian distributed; therefore, a nonparametric Friedman test for paired data was used to compare the kappa values obtained with ANIMAL and SyN and with the patch-based method. A p-value of less than 0.05 was deemed significant. When a significant difference was detected by the Friedman test, a nonparametric post hoc multicomparison test was used to determine which pair of methods was significantly different.

Results

The results of intra- and inter-rater variability of manual segmentation are presented in Table 2. The results of the center-of-gravity values between the deep brain structures obtained with 2 registration methods and the patch-based method are presented in Table 3.

The paired-t test showed a significant difference between the SyN registration and patch-based method ($p = 0.01$) and the ANIMAL registration and patch-based method ($p = 0.002$). There was no difference between ANIMAL and SyN registration methods in terms of distances ($p = 0.13$). The

Table 2 Mean kappa values for comparison of segmentation of three cerebral structures by two experts, each segmentation made three times on the same patient with Parkinson's disease

	Inter-rater		Expert 1		Expert 2	
	Mean	SD	Mean	SD	Mean	SD
Putamen	0.77	0.04	0.83	0.03	0.84	0.02
SN	0.68	0.02	0.80	0.03	0.79	0.02
STN	0.55	0.08	0.71	0.07	0.74	0.02

SN substantia nigra; *STN* subthalamic nucleus; *SD* standard deviation

Table 3 Mean distances and SD (in mm) between the center-of-gravity positions between manually segmented deep brain structures, and the automatically segmented labels from the 3 different methods: ANIMAL, SyN, and the patch-based

	Distance	
	Mean	SD
ANIMAL	1.51	0.88
SyN	1.48	0.92
Patch-based method	1.29	0.76

SD standard deviation

mean distances measured between the repeated segmentations of the two experts were 0.82 (SD 0.49) and 0.96 (SD 0.36) for Claire Haegelen and Louis Collins, respectively. These mean distances were not significantly different ($p = 0.15$).

The kappa index values obtained with the three methods are presented in Table 4 and Fig. 5.

ANIMAL and SyN-based method: The median kappa values computed between the labels obtained with ANIMAL and the manual labels were greater than 0.73 for the bilateral thalami, putamen, amygdala, hippocampus, and red nuclei, and between 0.62 and 0.67 for the bilateral substantia nigra and subthalamic nuclei (Fig. 5).

Patch-based method: The median kappa values computed between the labels obtained with the patch-based method and the manual labels were greater than 0.76 for the bilateral thalami, putamen, amygdala, hippocampus, and red nuclei, and between 0.62 and 0.70 for the bilateral substantia nigra and subthalamic nuclei (Fig. 5).

The Friedman test showed a significant difference between the three methods for the left amygdala ($p = 0.014$), left red nucleus ($p = 0.045$), and left ($p = 0.045$) and right thalami ($p = 0.001$) (Fig. 6). For the left amygdala, registration with the patch-based segmentation method was better than with ANIMAL. For the left red nucleus and left thalamus, registration with ANIMAL was better than with SyN. For the right thalamus, registration with ANIMAL was better than with the patch-based method.

Table 4 Dice-kappa values for three methods of segmentation of 14 cerebral structures in 10 patients with Parkinson's disease

	l_AG	r_AG	l_HC	r_HC	l_put	r_put	l_thal	r_thal	l_RN	r_RN	l_SN	r_SN	l_STN	r_STN
ANIMAL														
Mean	0.760	0.759	0.763	0.743	0.815	0.809	0.864	0.868	0.796	0.793	0.689	0.657	0.641	0.640
Median	0.773	0.738	0.792	0.760	0.814	0.809	0.854	0.858	0.781	0.774	0.671	0.640	0.624	0.638
Variance	0.0016	0.0007	0.002	0.0012	0.0012	0.0014	0.0008	0.0007	0.0027	0.002	0.0015	0.0033	0.004	0.006
SyN														
Mean	0.776	0.760	0.774	0.760	0.816	0.809	0.854	0.858	0.781	0.782	0.671	0.640	0.626	0.64
Median	0.776	0.759	0.792	0.771	0.814	0.809	0.854	0.867	0.792	0.782	0.671	0.640	0.624	0.65
Variance	0.0013	0.0005	0.003	0.0013	0.0009	0.0013	0.0008	0.0008	0.0030	0.0029	0.0009	0.002	0.003	0.006
Patch-based														
Mean	0.780	0.739	0.764	0.767	0.825	0.795	0.850	0.828	0.790	0.783	0.696	0.653	0.631	0.575
Median	0.810	0.761	0.780	0.766	0.822	0.790	0.851	0.819	0.785	0.784	0.704	0.675	0.621	0.626
Variance	0.008	0.005	0.0002	0.0001	0.005	0.0006	0.0006	0.0008	0.002 [a]	0.003[a]	0.003[a]	0.009[a]	0.005[a]	0.022[a]

l left; *r* right; *AG* amygdala; *HC* hippocampus; *put* putamen; *thal* thalamus; *RN* red nucleus; *SN* substantia nigra; *STN* subthalamic nucleus

[a] Values according to the automatic method only on the T2w analysis

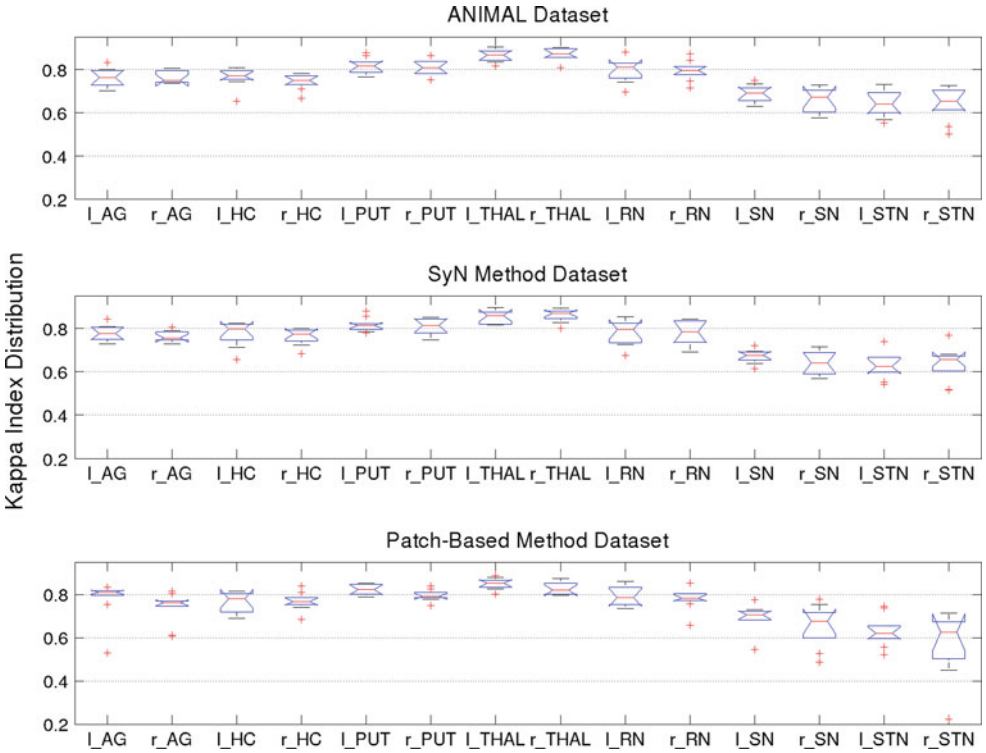

Fig. 5 Box plots of the kappa index distribution for ANIMAL registration (*top row*), SyN registration (*second row*), and patch-based method of segmentation (*bottom row*). *Boxes* represent the lower quartile, median (*red line*), and upper quartile of the kappa index distribution. Whiskers show the most extreme values within 1.5 times the interquartile range. Outliers (*red plus*) are data with values beyond the ends of the whiskers. l_AG: left amygdala; r_AG: right amygdala; l_HC: left hippocampus; r_HC: right hippocampus; l_PUT: left putamen; r_PUT: right putamen; l_THAL: left thalamus; r_THAL: right thalamus; l_RN: left red nucleus; r_RN: right red nucleus; l_SN: left substantia nigra; r_SN: right substantia nigra; l_STN: left subthalamic nucleus; r_STN: right subthalamic nucleus

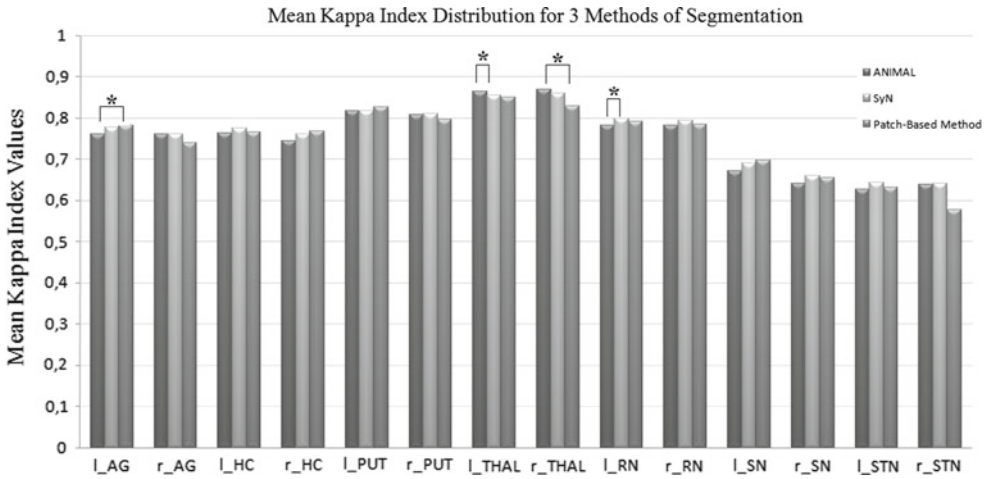

Fig. 6 Mean kappa index distribution for segmentation of cerebral structures in 10 patients with Parkinson's disease using ANIMAL registration, SyN registration, and patch-based method. *asterisk* represents a significant difference between two methods according to Friedman and post hoc multicomparison test (see the text in "Results" for details)

Discussion

The template-based segmentation approach is one method that has been developed to assist neurosurgeons in the target planning step of deep brain stimulation procedures. For example, Bardinet et al. [6] and Chakravarty et al. [7] developed different 3D histological atlases that were aligned to a MR image of the same brain obtained after autopsy, and to an

average MR image, respectively. The template can then be registered to the patient's MR image to delineate deep brain structures like the STN or GPm that will be targeted during surgery. We built a high-resolution/high-signal-to-noise ratio template by averaging the nonlinearly aligned MR images of many patients with Parkinson's disease. This process allowed us to obtain T1w and T2w templates with a high degree of anatomical detail for deep brain structures and the cerebral cortex. In contrast to the atlases of Bardinet et al. [6] and Chakravarty et al. [7], we have a direct high-quality visualization of the basal ganglia on the PD_T1_57 template and even more so on the PD_T2_42 template for those nuclei that are only visible on the T2w images, such as the red nucleus, substantia nigra, and subthalamic nucleus (see Fig. 1). The segmentation of the basal ganglia and deep brain structures is an important step in obtaining their 3D representation and importing them onto the patient's MR image. The manual segmentation of deep brain structures by the first author, Claire Haegelen, was found to be easier on the PD_T1_57 and PD_T2_42 templates than on a 3T digitized mono-subject template [8] obtained from a normal subject, but this fact was difficult to quantify and demonstrate, and beyond the scope of this paper where the goal was to compare three segmentation procedures. One of the limits of our study was that the segmentation used during methods' comparison was performed by only one expert. However, the segmentation quality was verified by a second expert, Xavier Morandi, and two experts (Claire Heagelen and D. Louis Collins) repeatedly segmented three cerebral structures of a patient with Parkinson's disease to estimate intra- and inter-rater reliability (see Table 2). The intra-rater segmentation reliability was quite high with values above 0.83 for putamen, around 0.80 for substantia nigra, and between 0.71 and 0.74 for the subthalamic nucleus. The inter-rater segmentation variability was slightly higher than the intra-rater for putamen and SN, while significantly higher for STN. This suggests that the manual delineation of STN from T1w MR scan is a very challenging task due to low contrast and small structure size that limits the possible consensus between expert's segmentations. Moreover, the manual segmentation of putamen, SN, and STN in patients with Parkinson's disease is difficult because these patients have a specific pathological loss of neurons of these three structures affecting their MRI visualization. In the literature, very few studies have done manual segmentation of basal ganglia from MRI or on histological slides [6,7] because of the great anatomical variability and the lack of very specific MRI sequences to identify accurately the STN, SN, and putamen in patients with basal ganglia pathological changes like those seen in Parkinson's disease. In the literature, many segmentation methods [18,29,34,36] are first tested on lateral ventricles, hippocampus, or amygdala, structures for which the boundaries and limits are better defined. In our opinion, these facts explain why in our study the STN had higher inter-rater variability compared to other reports of automatic segmentation techniques. Considering these manual segmentation difficulties, our results should be taken with the caveat that the inter-rater variability is relatively high. Nevertheless, the mean distances between centers of gravity were not different in the two sets of manual segmentations, and these distances were lower in labels obtained with the patch-based method than with the two registration methods. Furthermore, the mean Dice-kappa values for inter-rater segmentation are at approximately the same level as the mean Dice-kappa values comparing the manual labels and those generated by the three automatic techniques. In future, we will investigate the effect of consensus segmentations from multiple experts.

To our knowledge, this is the first study to compare different registration approaches and the patch-based method to identify deep brain structures in the context of Parkinson's disease. We applied two nonlinear registration-based segmentation tools to MR images, as nonlinear registration has been shown to perform better than linear registration for the automated identification of subcortical targets [30–32]. The accuracy of the automated segmentation depends on the accuracy of registration. In the study by Klein et al. [33], the two nonlinear tools, ANIMAL and SyN, were assessed based on overlap measures of manually labeled anatomical regions, as was done in our study. SyN was one of the methods that reached top rank for all label sets [33]. In our case, the two nonlinear registrations were quite similar for all the structures, even for the smallest ones, except for the left thalamus and red nucleus, for which ANIMAL performed better than SyN. The advantage of this type of segmentation approach is that the manual segmentation is performed on a high-contrast image (i.e., the template), which facilitates the delineation procedure. However, the difficulty is then in registering the template with an individual MR image because of inter-subject variability. A linear (or affine) transformation consists in transforming a template into the image space of the patient's MR data. Although it can account for variations in size, orientation, and position, a linear transformation cannot radically alter the shape of the template as is needed to account for anatomical differences between the patient's image and the template. Instead, nonlinear transformations are needed to address this issue. By using nonlinear transformations, we obtained a more accurate segmentation result; however, the nonlinear registration of the template to the patient MR image is time-consuming, taking in our case from 30 min to 1 h per patient for ANIMAL and from 1 to 3 h per patient for SyN.

Another way to obtain better segmentation accuracy than by registering images to a single labeled template is to fuse several labeled images [34]. Fusing the labeled images in the training library can be done by nonlinear registration [34, 35] or patch comparison [18]. The advantage of label fusion approaches is that they can handle inter-subject variability

correctly by aggregating labels from several images. In this manner, different anatomical patterns can be optimally fused to fit the patient's anatomy. The patch-based label fusion approach evaluated here is also less time-consuming than nonlinear registration, since it required only 1 min of processing.

The difficulty with the patch-based label fusion approach is that it requires manual segmentation of structures on the MR images of several subjects to form the library. Not only is this task time-consuming, but it is also difficult due to low-contrast image, especially in the context of Parkinson's disease where motion artifacts greatly impact the image quality. While the overall segmentation quality was similar for the patch-based and registration-based methods, there was more variability in the segmentation quality of the patch-based approach (see Fig. 5, Table 4). As shown in the study by Coupe et al. [18], the quality of the segmentation of patch-based label fusion depends on the number of training templates in the library. During our leave-one-out experiment, only nine templates were used to segment the subject under study. This number of training templates is small compared with what is suggested by Coupe et al. [18] (between 20 and 30). By increasing the number of training templates, the segmentation quality is expected to increase and the robustness of the segmentation is expected to improve.

The kappa metric is often used to compare the agreement between structures in the literature [18,32,34–36]. The kappa metric depends on the surface-to-volume ratio of the structure, with decreasing kappa values in small structures for which the ratio increases. We found lower kappa values for the hippocampus than what is reported by Coupe et al. [18] ($\kappa = 0.884$) and Collins et al. [34] ($\kappa = 0.887$), and also for the amygdala, for which Collins et al. [34] found $\kappa = 0.826$. As mentioned, this lower segmentation quality may be due to the smaller number of training subjects used in our segmentation procedure. It might also be due to the low-contrast images of patients with Parkinson's disease or to the use of a different database [18,34]. Patenaude et al. [36] used a Bayesian appearance model approach to provide accurate segmentation of 15 subcortical structures on 336 manually labeled T1w MR images. They found that the best structures in terms of kappa values were the thalamus and putamen with a median kappa as high as 0.85, and the hippocampus with a median kappa between 0.80 and 0.85, as in our study, though we found lowest kappa values for the same structures. Patenaude et al. [36] found that the worst structures in terms of kappa values were the amygdala and the nucleus accumbens because, in those cases, the kappa penalized small structures or structures with a higher surface-to-volume ratio. We also found lower median kappa values for the smallest structures, the STN and substantia nigra, but not for the red nucleus. One hypothesis may be that, in patients with Parkinson's disease, compared with the red nucleus, the STN and substantia nigra are more affected by the loss of dopaminergic neurons and therefore more difficult to identify with automated segmentation tools.

The three methods were relatively equivalent in terms of mean and median kappa values, suggesting they perform similarly in producing the automated segmentation of deep brain structures. Differences between the three methods were observed only for some of the structures, which mean there was, in fact, very little dissimilarity. However, when applied to a new patient, the patch-based method is notably more efficient than the other two methods, as it avoids the lengthy time-consuming, nonlinear registration step to achieve segmentation. Moreover, considering the resulting high inter-rater variability in our study, the automatic segmentation was as variable as another expert rater but was not subject to intrarater variability. The patch-based method probably underestimates the size of the labels because of the mean center-of-gravity values between the labels lower than with the two other registration methods. The STN deep brain stimulation proved a great efficacy in reducing the symptoms of Parkinson's disease [37,38]. The patch-based segmentation method may improve the surgical workflow by quickly and automatically identifying the STN on the preoperative MRI data of new patients, thus the reducing part of the preoperative planning time in STN targeting. In clinical practice, the patch-based method is not yet applicable because it requires an engineer. From a point of view of a neurosurgeon, the three methods of segmentation were accurate enough to visualize the STN or the thalamus and to be applicable to new patients. The limits were the long time-consuming step of the ANIMAL and SyN registrations method in clinical practice. More work is needed to have a software solution that is usable by the clinicians without the supervision of the engineers.

Conclusions

We have developed the first step toward a multimodal database on deep brain stimulation, specifically the construction of an MR image template specific to Parkinson's disease, and also evaluated a method of achieving the accurate segmentation of the basal ganglia and deep brain structures on patient data.

We compared three methods to determine the best strategy for segmenting the basal ganglia and deep brain structures on patients' MR images. We assessed that the intrarater variability between the manual segmentations was lower than the inter-rater variability. Although template-based approaches are the most frequently used techniques to warp basal ganglia segmentation onto MR images, we found that the patch-based method provides similar results and is much less time-consuming.

Acknowledgments The authors thank Amanda Leonard for her help with the English language. The authors thank the Société Française de Neurochirurgie, Baxter Society and Medtronic for their support of this study.

Conflict of interest None.

References

1. Benabid AL, Chabardes S, Mitrofanis J, Pollak P (2009) Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. Lancet Neurol 8:67–81
2. Koller W, Pahwa R, Lyons K, Wilkinson S (2000) Deep brain stimulation of the Vim nucleus of the thalamus for the treatment of tremor. Neurology 55:S29–S33
3. Vayssiere N, Hemm S, Zanca M, Picot MC, Bonafe A, Cif L, Frerebeau P, Coubes P (2000) Magnetic resonance imaging stereotactic target localization for deep brain stimulation in dystonic children. J Neurosurg 93:784–790
4. Parsons TD, Rogers SA, Braayen AJ, Woods SP, Troster AI (2006) Cognitive sequelae of subthalamic nucleus deep brain stimulation in Parkinson's disease: a meta-analysis. Lancet Neurol 5:578–588
5. York MK, Wilde EA, Simpson R, Jankovic J (2009) Relationship between neuropsychological outcome and DBS surgical trajectory and electrode location. J Neurol Sci 287:159–171
6. Bardinet E, Bhattacharjee M, Dormont D, Pidoux B, Malandain G, Schüpbach M, Ayache N, Cornu P, Agid Y, Yelnik J (2009) A three-dimensional histological atlas of the human basal ganglia. II. Atlas deformation strategy and evaluation in deep brain stimulation for Parkinson disease. J Neurosurg 110:208–219
7. Chakravarty MM, Bertrand G, Hodge CP, Sadikot AF, Collins DL (2006) The creation of a brain atlas for image guided neurosurgery using serial histological data. Neuroimage 30:359–376
8. Lalys F, Haegelen C, Ferre JC, El-Ganaoui O, Jannin P (2010) Construction and assessment of a 3T-MRI brain template. Neuroimage 49:345–354
9. Schaltenbrand G, Wahren W (1977) Atlas for stereotaxy of the human brain. Thieme, Stuttgart
10. Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Georg Thieme Verlag, Stuttgart
11. Yelnik J, Bardinet E, Dormont D, Malandain G, Ourselin S, Tandé D, Karachi K, Ayache N, Cornu P, Agid Y (2007) A three-dimensional, histological and deformable atlas of the human basal ganglia. I. Atlas construction based on immunohistochemical and MRI data. Neuroimage 34:618–638
12. D'Haese PF, Pallavaram S, Li R, Remple MS, Kao C, Neimat JS, Konrad PE, Dawant BM (2010) CranialVault and its CRAVE tools: a clinical computer assistance system for deep brain stimulation (DBS) therapy. Med Image Anal. doi:10.1016/j.media.2010.07.009
13. Guo T, Finnis KW, Parrent AG, Peters TM (2006) Visualization and navigation system development and application for stereotactic deep-brain neurosurgeries. Comput Aided Surg 11:231–239
14. Guo T, Parrent AG, Peters TM (2007) Surgical targeting accuracy analysis of six methods for subthalamic nucleus deep brain stimulation. Comput Aided Surg 12:325–334
15. Collins DL, Holmes CJ, Peters T, Evans AC (1995) Automatic 3-D model based neuroanatomical segmentation. Hum Brain Mapp 3:190–208
16. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, Brain Development Cooperative Group (2011) Unbiased average age-appropriate atlases for pediatric studies. Neuroimage 54:313–327
17. Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12(1):26–41
18. Coupé P, Manjon JV, Fonov V, Pruessner J, Robles M, Collins DL (2011) Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. Neuroimage 54:940–954
19. Neelin P (1998) The MINC file format:from bytes to brains. NeuroImage 7(4):786
20. Sled JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17:87–97
21. Nyul LG, Udupa JK, Saha PK (2003) Incorporating a measure of local scale in voxel-based 3-D image registration. IEEE Trans Med Imaging 22:228–237
22. Wells WMIII, Viola P, Atsumi H, Nakajima S, Kikinis R (1994) Multi-modal volume registration by maximization of mutual information. Med Image Anal 1(1):35–51
23. Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18:192–205
24. Smith SM (1998) Fast robust automated brain extraction. Hum Brain Mapp 17:143–155
25. Mai JK, Paxinos G, Voss T (2008) Atlas of the human brain. Elsevier Inc, Amsterdam
26. Coupé P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C (2008) An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans Med Imaging 27:425–441
27. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC (1994) Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Imaging 13:716–724
28. Dice L (1945) Measure of the amount of ecological association between species. Ecology 26(3):297–302
29. Hu S, Coupe P, Pruessner JC, Collins DL (2011) Appearance-based modeling for segmentation of hippocampus and amygdala using multi-contrast MR imaging. Neuroimage 58:549–559
30. Bhattachargee M, Pitiot A, Roche A, Dormont D, Bardinet E (2008) Anatomy-preserving nonlinear registration of deep brain ROIs using confidence-based block-matching. In: LNCS MICCAI, part II, Springer, vol 5242, pp 956–963. doi:10.1007/978-3-540-85990-1_115
31. Bardinet E, Dormont D, Malandain G, Bhattachargee M, Pidoux B, Saleh C, Cornu P, Ayache N, Agid Y, Yelnik J (2005) Retrospective cross-evaluation of an histological and deformable 3D atlas of the basal ganglia on series of Parkinsonian patients treated by deep brain stimulation. In: LNCS MICCAI, Springer, vol 3750, pp 385–393. doi:10.11007/11566489_48
32. Chakravarty MM, Sadikot AF, Germann J, Hellier P, Bertrand G, Collins DL (2009) Comparison of piece-wise linear, linear, and nonlinear atlas-to-patient warping techniques: analysis of the labelling of subcortical nuclei for functional neurosurgical applications. Hum Brain Mapp 30:3574–3595
33. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46:786–802
34. Collins DL, Pruessner JC (2010) Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. Neuroimage 52:1355–1366

35. Aljabar P, Heckeman RA, Hammers A, Hajnal JV, Rueckert D (2009) Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46:726–738
36. Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011) A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 56:907–922
37. Antonini A, Isaias IU, Rodolfi G, Landi A, Natuzzi F, Siri C, Pezzoli G (2011) A 5-year prospective assessment of advanced Parkinson disease patients treated with subcutaneous apomorphine infusion or deep brain stimulation. J Neurol 258: 579–585. doi:10.1007/s00415-010-5793-z
38. Merola A, Zibettu M, Angrisano S, Rizzi L, Lanotte M, Lopiano L (2011) Comparison of subthalamic nucleus deep brain stimulation and duodopa in the treatment of advanced Parkinson's disease. Mov Disord 26(4):664–670