A Cross Validation Study of Deep Brain Stimulation Targeting: From Experts to Atlas-Based, Segmentation-Based and Automatic Registration Algorithms

F. Javier Sanchez Castro*, Claudio Pollo, Reto Meuli, Philippe Maeder, Olivier Cuisenaire, Meritxell Bach Cuadra, Jean-Guy Villemure, and Jean-Philippe Thiran

Abstract—Validation of image registration algorithms is a difficult task and open-ended problem, usually application-dependent. In this paper, we focus on deep brain stimulation (DBS) targeting for the treatment of movement disorders like Parkinson's disease and essential tremor. DBS involves implantation of an electrode deep inside the brain to electrically stimulate specific areas shutting down the disease's symptoms. The subthalamic nucleus (STN) has turned out to be the optimal target for this kind of surgery. Unfortunately, the STN is in general not clearly distinguishable in common medical imaging modalities. Usual techniques to infer its location are the use of anatomical atlases and visible surrounding landmarks. Surgeons have to adjust the electrode intraoperatively using electrophysiological recordings and macrostimulation tests. We constructed a ground truth derived from specific patients whose STNs are clearly visible on magnetic resonance (MR) T2-weighted images. A patient is chosen as atlas both for the right and left sides. Then, by registering each patient with the atlas using different methods, several estimations of the STN location are obtained. Two studies are driven using our proposed validation scheme. First, a comparison between different atlas-based and nonrigid registration algorithms with a evaluation of their performance and usability to locate the STN automatically. Second, a study of which visible surrounding structures influence the STN location. The two studies are cross validated between them and against expert's variability. Using this scheme, we evaluated the expert's ability against the estimation error provided by the tested algorithms and we demonstrated that automatic STN targeting is possible and as accurate as the expert-driven techniques currently used. We also show which structures have to be taken into account to accurately estimate the STN location.

*F. J. Sanchez Castro is with the Signal Processing Institute, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

C. Pollo and J.-G. Villemure are with the Department of Neurosurgery, Centre Hospitalier Universitaire Vaudois (CHUV), CH-1011 Lausanne, Switzerland.

R. Meuli and P. Maeder are with the Department of Radiology of the Centre Hospitalier Universitaire Vaudois (CHUV), CH-1011 Lausanne, Switzerland. (e-mail: francisco.sanchezcastro@epfl.ch).

O. Cuisenaire, M. Bach Cuadra, and J.-P. Thiran are with the Signal Processing Institute, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

Digital Object Identifier 10.1109/TMI.2006.882129

Index Terms—Atlas-based segmentation, B-splines, deep brain stimulation, demons, nonrigid registration, preoperative targeting, surgical planning, validation.

I. INTRODUCTION

IMAGE registration has become a common and useful tool in the medical environment. It can be defined as the problem of finding an optimal geometric transformation between corresponding image data. Its aim is to compute the spatial transformation which maps each point of one image onto its corresponding point of another image. Registration methods are widely used for helping clinicians because of the difficulty of fusing images mentally, even more so when they come from different patients and/or different modalities. Typical clinical and surgical applications include preprocedural planning and simulation, interventional radiology, diagnostic radiology, minimally invasive procedures, radiation therapy, intraoperative navigation, robot-assisted interventions, etc. Nowadays, rigid registration is a well-covered domain [1] with efficient solutions for the most common applications. Current research mostly focuses on nonrigid registration methods. Many different approaches can be found in the literature [2]–[4]. Nevertheless, there is not a universal solution due to the complex framework of nonrigid registration. This framework can be divided and classified following different criteria [2]: the *dimensionality* of the data involved, the *feature space* to base the registration on, the *kind and domain of the transformation model*, the *modalities* of the data and the *subjects* involved, the *user interaction*, the *similarity metric* to assess the matching quantitatively, the *search space* in which the optimal solution has to be found, and the *search strategy* including optimization of the objective function and implementation details. Together with this extremely complex framework, the validation of the results provided by different techniques and for different applications is a crucial issue but remains an unsolved problem. The lack of gold standards and the usual unavailability of a ground truth makes this problem difficult to solve. Some valuable contributions have nevertheless been made in this area. In [5], a reference is estimated from a collection of segmentations for atlas-based segmentation purposes. In [6], an evaluation of different nonrigid registration

methods is performed using global and local measures focusing on the matching of cortical areas. In [7], nonrigid registration is applied over a labeled brain atlas to segment gross anatomical structures. However, validation assessments cannot be generalized to all the applications because of their different and specific requirements. Validation of medical image registration is closer to an application-dependent problem.

In this work, we focus on deep brain stimulation (DBS) targeting, which is a critical step in the surgical treatment of a number of neurological diseases like Parkinson's disease (PD) and essential tremor. Accurate preoperative targeting influences directly and critically the operating time and, most importantly, the outcome of the surgery. Before DBS was proposed for the treatment of PD [8], medications dominated with the introduction of *levodopa*. Unfortunately, this medical treatment has important shortcomings (drug-induced effects) and benefits decrease with time. These problems, together with the increase of basal ganglia circuitry knowledge, have led to the development of surgical-based techniques. Two kinds of approaches can be distinguished: lesion-based (thalamotomy and pallidotomy) techniques and DBS. The first one has shown a good performance in treating tremor in one side of the body and reducing dyskinesias (involuntary twisting and writhing movements) but cannot be applied bilaterally because this may produce severe undesirable and permanent effects on speech, behavior, or cognition. On the other hand, DBS can be performed safely in both sides of the brain and is reversible. DBS involves implantation of an electrode deep inside the brain. This electrode delivers electric current to specific brain cells shutting down parkinsonian symptoms. The subthalamic nucleus (STN) [9] has turned out to be, based on wide surgical experience, the optimal target for DBS [10], [11]. The STN is a small biconvex lens-shaped structure of diameter ranging from 4 to 10 mm. Unfortunately, the STN is in general not clearly delineated on magnetic resonance (MR) images.

The protocol used for STN target selection varies among different clinical centers. Common strategies are the use of stereotactic anatomical brain atlases [12], [13] (where usually the target given by the atlas is corrected by the expert based on the width of the third ventricle) and the use of visible surrounding anatomical landmarks [14]. In our state-of-the-art protocol, a DBS procedure begins by fixing the stereotactic head frame to the patient's skull, which is used as a coordinate reference system. Next, two kinds of imaging studies are acquired preoperatively to target and plan the electrode placement: MR T1-weighted and MR T2-weighted. After identification of the anterior commissure (AC), posterior commissure (PC) and midcommissural point (MCP) on a three-dimensional T1-weighted sequence, inversion recovery (IR) T2-weighted coronal slices are performed orthogonally to the AC-PC line. On the slice showing the anterior pole of the red nucleus (RN), the target is placed in the inferolateral portion of the subthalamic zone, limited superiorly by the thalamus, laterally by the internal capsule, inferiorly by the substantia nigra, and medially by the midline [14]. Then, the target coordinates are reported onto the T1-weighted image space where trajectories are planned. Following these preoperative trajectories, a small hole is drilled in the patient's skull, which is fixed to the operating table, and

the electrode is introduced until the preoperative STN target is reached. This location is only an estimation of the real STN location. Therefore, during the progression of the electrode, surgeons may need to correct the preoperative estimation using both microelectrophysiological recordings and macro-stimulation tests. Throughout the electrophysiological recordings, neuronal activity can be recorded and interpreted to identify STN patterns as well as to note the side effect of stimulation and the difference between the voltage that induces the effect and the voltage that induces side effects. Furthermore, the observed patient's reactions like disappearance of tremor or diminution of rigidity (the patient is only under local anesthesia) helps the placing of the electrode. Nevertheless, different locations along the same track can produce similar responses. Moreover, a guide of parallel tracks (usually five, separated by 2 mm) allowing to change the initial choice, if it is not satisfying, brings the accurate preoperative target selection of crucial importance.

The main goals of the validation study presented in this paper are the following. First, the construction of a ground truth for the position of the STN regarding the preoperative targeting. Second, the study of the intraexpert and interexpert variability in identifying the STN location visually using the procedure described below. Third, to evaluate the possibility of automatically locating the STN by atlas-based segmentation using existing registration techniques and to compare their performance and usability. We focus also on answering the following questions. Which combination of feature-structures most influences the STN location estimation through a registration process? Does the use of these structures alone produce an accurate estimation of the STN target? How well does the structure-based registration perform compared to the whole-brain nonrigid registration algorithms and atlas-based methods? Are there any structures or combination of structures that produces an error in estimating the STN location comparable to the expert variability when targeting clearly visible STNs?

To reach the goals and answer these questions, a ground truth is constructed directly from clearly visible STNs on T2-weighted images. Then one patient volume is chosen as an atlas and each of the remaining patients mapped back to the atlas using various atlas-based, segmentation-based, and automatic registration methods. An estimation of the STN location is obtained and compared to the ground truth. We demonstrate that using accurate registration algorithms, automatic STN localization is not only possible but is as accurate as expert-based methods currently in use. We also study the influence of surrounding structures in the STN estimation. Some efforts for automatically estimating the STN target have been reported recently [15]–[18]. They construct the ground truth through indirect methods by registering anatomical, histological, and/or electrophysiological atlases with the reference patient, model or template but, as far as we know, our study is the first one that constructs the ground truth in a direct way from visible STNs and that directly studies the influence of neighboring structures in the estimation of the STN's location.

This paper is organized as follows. First, we give a description of the data set considered in this study. Then, the protocol used to construct the ground truth and the validation scheme procedure are described. Next, the atlas-based, automatic nonrigid
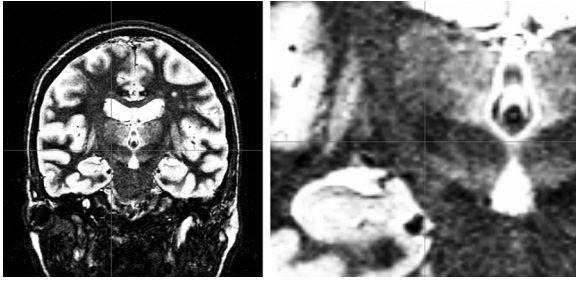
Fig. 1. T2-weighted image of a patient with clearly visible STN (zoomed in region of interest on the right panel).

and segmentation-based methods under investigation are presented. In Section III, we first analyze the expert's variability and then we study the effectiveness of the methods in three steps. First, the atlas-based and automatic nonrigid registration algorithms are tested against the expert's variability. Second, a study of the neighboring structures which influence the STN location is performed. Third, a comparison and validation of all the methods together is presented. Finally, we discuss the results and report the conclusions.

II. MATERIAL AND METHODS

A. Data

In this study, a set of 39 bilaterally implanted parkinsonian patients (78 STNs) were considered as a starting database. Two kinds of images were acquired preoperatively for each patient: a three-dimensional (3-D) T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) MRI sequence (Siemens Vision, 1.5T, Erlangen, Germany) TR 9.7 ms, TE 4 ms, number of slices/slice thickness: 164/1.40 mm, FOV 280 $\times$ 280, matrix 256 $\times$ 256, pixel size 1.09 $\times$ 1.09 mm, and few coronal slices (due to the acquisition time required for this kind of imaging sequence) of an inversion recovery (IR) T2-weighted, TR 2560 ms, TE 4 ms, number of slices/slice thickness: 7/3 mm, FOV 300 $\times$ 300, matrix 512 $\times$ 512, pixel size 0.59 $\times$ 0.59 mm. At least one series of IR T2-weighted slices was performed orthogonal to the AC–PC line in order to localize the anterior pole of the red nucleus. On the red nucleus slice, we identified the subthalamic zone. Taking profit from the fact that in some specific patients the STN is visible in MR T2-weighted images, a reference was constructed and used as a ground truth. To do this, neurosurgeons and radiologists were asked to select patients with a clearly visible STN in MR T2-weighted images from our patient database. After exhaustive inspection, eight subjects were selected (16 STNs). Only these images were used in the remainder of this work. In Fig. 1, one can see a T2-weighted image of a patient with a clearly visible STN.

B. Ground Truth Construction

Two experts were selected to generate the ground truth: a neurosurgeon and a radiologist, both with a large experience in PD surgery and targeting. They were asked separately to select the STN target points for each patient of the selected T2 series. This procedure was repeated five times on different days to avoid the experts being influenced by previous targeting choices. Then, a paired t-test of the hypothesis that the target coordinates that the experts selected came from distributions with equal means was performed over each coordinate (x, y, z) at a 1% significance level. Based on the results, this hypothesis could not be rejected. Therefore, for our reference, we used the geometrical mean of the two cloud of points given by each expert as *real targets*. The real targets were collected for the eight patients involved in this study and with these data we were able to compute intraexpert repeatability and interexpert variability, as shown in Section III-A. Finally, among the eight selected patients, the experts selected the ones with the most clearly visible STN as reference or atlas subjects, one for the right STN and one for the left STN. These two patients have important anatomical differences (e.g., width of the ventricles), as can be seen in Fig. 4, which will be useful to compare the performance of the methods for large anatomical differences.

C. Validation Scheme

The first step of our validation procedure consists of projecting the T2 real target coordinates into the T1 space where we dispose of whole-brain images. To do this, intrapatient T1-T2 rigid (translation and rotation) registration is performed for each patient [19]. Then, by nonrigidly registering each patient with the atlas a projection of the patient's STN is obtained, which is an estimation of the STN location given by the registration technique considered. This estimation is obtained by registering the MR T1-weighted images of the patients (leaving out the reference) with the atlas. Finally, Euclidean distances from estimated to real targets are computed and then statistics are derived in order to compare and validate the performance of the different tested methods as well as the influence of different surrounding structures in the STN location. A flowchart of this validation scheme is shown in Fig. 2.

D. Automatic Registration and Atlas-Based Methods

In this comparison-validation study, an atlas-based technique widely used in a medical environment, as well as three different nonrigid registration algorithms selected for their wide use in medical image processing were tested using whole brain images. The nonrigid registration methods were implemented using the ITS-EPFL medical image processing library [20]. These methods are the following as follows.

- Atlas-based (AC–PC) targeting. The use of anatomical brain atlases [12], [13] together with the AC-PC referential is a common procedure used for DBS targeting purposes mainly when the STN is not clearly visible in MR T2-weighted images, which is the usual case. Experts have to locate the anterior and posterior commissures (AC–PC points) in the MR T1-weighted images. In this work, we follow the Schaltenbrand–Wahren protocol [12]: using this stereotaxic atlas of the brain and taking as the origin of coordinates the midcommissural point (MCP, middle point of the AC–PC line), this technique estimates that the STNs are located at coordinates: anteroposterior (AP) −3 mm, lateral (LAT) ±12 mm and vertical (VERT) −4 mm.
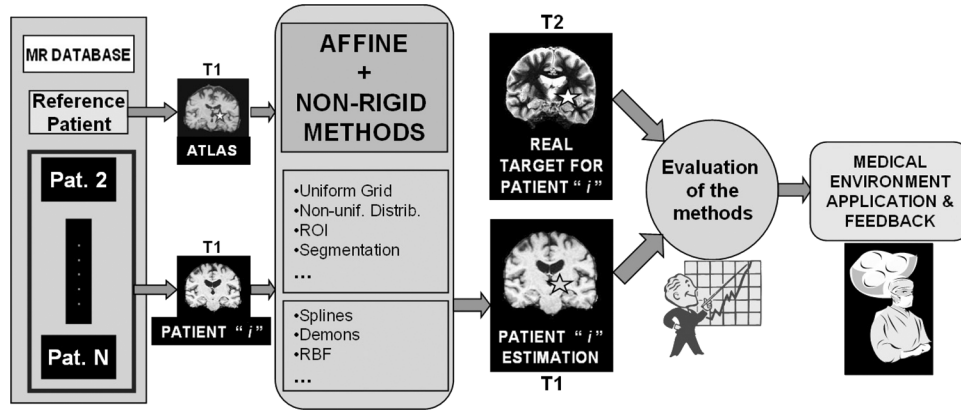
Fig. 2. Validation scheme for target estimation using different methods.

- Affine registration. This is an independent implementation based on the work of Maes *et al.* [19]. The mutual information between the images to be registered is maximized [21], [22] by optimizing the 12 degrees-of-freedom (translation, rotation, scaling and shearing). This affine transformation can be written as

$$T(x,y,z) = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \theta_{14} \\ \theta_{24} \\ \theta_{34} \end{pmatrix}$$

where the degrees-of-freedom are parameterized by the coefficients θ. This is performed using a coarse-to-fine multiscale approach and a two-step optimization. First, a global search using genetic algorithms [23] and next a local optimization using the steepest descent method [24] were used to find the best parameters. Affine registration is also used as a prealignment step for nonrigid transformations described below.

- Demons algorithm. This is an intensity-based algorithm proposed by Thirion [25], [26], based on the concept of optical flow. The problem of image matching is approached as a diffusion process, in which object boundaries in the reference image F are viewed as semi-permeable membranes. The other (so-called floating) image G is considered as a deformable grid, and diffuses through these interfaces driven by the action of effectors situated within the membranes. These effectors are also called *demons* by analogy with Maxwell's demons. In the case of voxel-by-voxel intensity similarity, the instantaneous displacement vector for each voxel is

$$\overrightarrow{d} = -\frac{(g-f)\overrightarrow{\nabla}f}{|\overrightarrow{\nabla}f|^2 + (g-f)^2}$$

where f and g are the intensity image of F and G, respectively, which are previously intensity-equalized. Thus, there is a displacement in the direction of the gradient provided by both a difference in image intensities and a reference image gradient different from zero. The deformation algorithm is applied by iterating in a hierarchical

coarse-to-fine multiscale way to make the algorithm more robust with regard to large differences (these differences could render optical flow methods completely ineffective because the assumption of a small displacement is violated). The smoothness of the displacement field is imposed by filtering with a Gaussian filter of standard deviation σ (*elasticity parameter*) chosen empirically [27]. In our case, the parameter σ was chosen by exhaustive search, between 0.6 and 2.0 mm by steps of 0.2 mm, minimizing the distance between the estimated STN and the real targets. Finally, σ of 1 mm was chosen. Also, since the above equation is asymmetrical, it gives different results depending on which image is chosen as the reference and which is chosen to be floating. As proposed in [26], consistency between the forward and the reverse deformation fields can be maintained through a bijective implementation. Both the deformation fields D_{12} (the deformation field warping image 1 onto image 2) and D_{21} (the deformation field warping image 2 onto image 1) are computed and the residual $R = D_{12} \circ D_{21}$ is distributed between these two fields.

- B-splines algorithm. This is a mutual information-based free-form deformation algorithm whose displacement field is modelled as a linear combination of B-splines lying on a regular grid (uniformly spaced control points) similar to the method proposed by Rueckert *et al.* in [28]. This deformation produces a C^2 continuous and smooth transformation. The deformation that maximizes the mutual information between the two images involved is computed at each grid point placed on the floating image. The transformation is propagated to the rest of the image using the standard B-spline expansion with cubic splines

$$d(x) = \sum_{k\in\mathbb{Z}} c(k)\beta^3(x-k)$$

where $\beta^3(x) = \beta^0 * \beta^0 * \beta^0 * \beta^0(x)$, $c(k)$ are the B-spline coefficients and β^0 a rectangular pulse. To speed up the optimization process, the algorithm was implemented through a hierarchical multiscale scheme (both for images and grid of control points) and using the communication utilities for distributed memory architectures using the
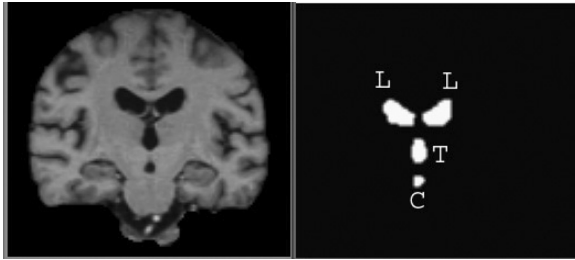
Fig. 3. Segmented structures used for feature-based registration: lateral ventricles (L), third ventricle (T), and interpeduncular cistern (C).

TABLE I
EXPERT VARIABILITY STATISTICS. PREOPERATIVE TARGETING OVER CLEARLY VISIBLE STNS

$mean \pm std$	Expert 1	Expert 2
intra	1.06 ± 0.61 mm	0.80 ± 0.52 mm
inter	1.61 ± 0.29 mm	1.40 ± 0.38 mm

TABLE II
INTRAEXPERT VARIABILITY TARGETING FOR THE REFERENCE STNS

$mean \pm std$	Left Ref.	Right Ref.
Expert 1	1.10 ± 0.32 mm	0.75 ± 0.38 mm
Expert 2	0.79 ± 0.30 mm	0.38 ± 0.25 mm

MPICH implementation of the message passing interface (MPI) [29]. The good interpolation properties and the suitability for multiscale processing of the B-splines are well known [30] and its deformability can be controlled by changing the spacing between the control points of the grid which we have set at 12 mm to be able to cope with highly local nonrigid deformations.

E. Segmentation-Based Registration

As the STN is located in an intensity homogeneous region in the MR T1-weighted images, the transformation given by the registration algorithms is driven by the deformations computed for the surrounding visible structures (e.g., the thalamus, lateral ventricles, third ventricle, putamen, red nucleus, globus pallidus, interpeduncular cistern, mesencephalon, internal capsule, etc.). These structures can be segmented. The question is whether by registering only these segmented structures or a particular subset of them and by applying the resulting deformation field to the original image, an accurate estimation of the STN location can be obtained. Yet, we wish to be able to target the STN automatically. Therefore, we have to focus on automatically-segmentable structures rather than complex ones that need manual delineations. Four easy-segmentable structures were chosen for this study: lateral (left and right) ventricles, third ventricle and interpeduncular cistern, which we denote as L, T, and C, respectively (see Fig. 3). As these structures are filled with cerebrospinal fluid (CSF), their contours on typical MR images can easily be delimited. The segmentation was done using a semiautomatic thresholding method but can be performed in a fully automatic way by using, for instance, *a priori* knowledge of shape [31]. The procedure is as follows. Each segmented structure or a combination of them (binary mask) of the atlas is registered with the corresponding segmented structure (binary mask) of all the patients using a tandem of affine-demons nonrigid registration (see Section II-D). Note that only the lateral ventricle corresponding to the STN of the same brain side is used for estimating its location. First, an affine registration using whole-brain images is performed to cope with global misalignment (as for the above B-splines or demons methods). Then, a demons registration is applied between the binary masks of the structures considered both for the atlas and the patient under study. The elasticity parameter σ is the same as for the whole brain images. Some advantages are obtained from the fact of using the demons algorithm with a binary mask. The computation time is decreased since only a small region of interest is registered. The intensity equaliza-

tion of the images is no longer needed. Furthermore, the demons points to be considered are only those placed on the surface of the segmented volumes, where the gradient is nonzero, reducing the number of calculations required.

III. RESULTS

A. Expert Targeting Variability

In order to evaluate the repeatability or intraexpert variability of the expert targeting, we computed the centroid of each cloud of STN points targeted by each expert and we calculated the Euclidean distances from the centroid to each of these expert points yielding a *centroid variability* measure. In Table I, for both experts these statistics are shown. These quantities allow us to get an idea of the surgeon variability and its accuracy when clicking over a clearly visible preoperative target. The interexpert variability calculated as the Euclidean distances from each expert click to the ground truth gives a mean and unbiased standard deviation shown in Table I. If we only consider the two STNs used as a reference we obtain a centroid (intraexpert) variability that is shown in Table II for the left and right sides, respectively, and for the different experts.

In Fig. 4, the points targeted by the experts for the left [Fig. 4(a)–(c)] and right [Fig. 4(d)–(f)] STNs chosen as a reference can be seen. In blue, we show the targeting of expert 1, in red the targeting of expert 2, and in black the mean point used as a reference. For visualization purposes, each point was projected onto the three orthogonal planes passing through the centroid and shown using a circle (radius of 1 mm).

We can also decompose these distances in anteroposterior (AP), lateral (LAT), and vertical (VERT) coordinates which allows a direct comparison of the mean STN location for our database with the usual STN coordinates from the atlases (e.g., [12]). In Table III, we show the mean and unbiased standard deviation of the coordinates referred to the MCP for the 16 STNs used in this study.

Considering that the experts are targeting a point within a volumetric structure, we can see from the results that the cloud of each targeted STN is spread in around 1.5 mm, corresponding approximately to the size of one voxel of the MR T1-weighted acquisition. Moreover, the mean coordinates of these targeted points show slight differences to typical anatomical atlas coordinates. This can be explained by the intentional placement of the target by experts at the inferolateral portion of the subthalamic nucleus [14] while typical atlas coordinates are taken at the volumetric center of the structure.
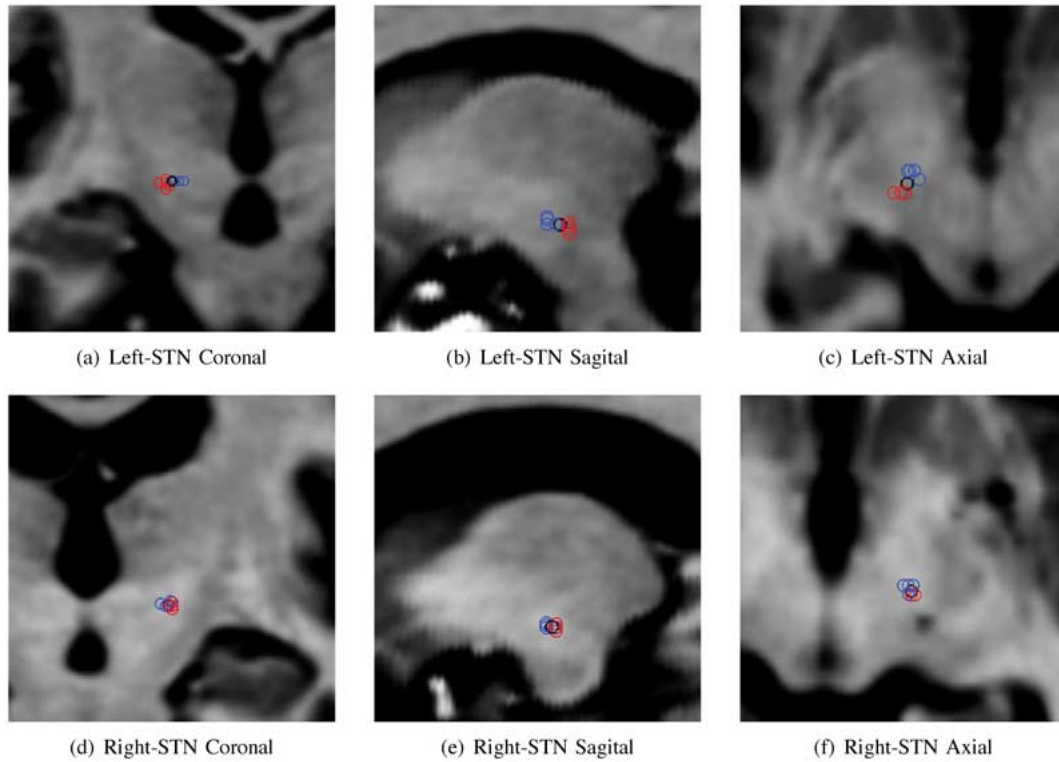
Fig. 4. Reference STN expert targeting. Expert 1 in red, expert 2 in blue, ground truth in black (each point represented by a circle of radius 1mm).

TABLE III
MEAN STN COORDINATES REFERRED TO MCP FOR OUR DATASET AND FOR THE SHALTENBRAND–WAHREN ATLAS

Coordinates	Dataset	S-W atlas
AP	-3.86 ± 0.94 mm	-3 mm
LAT	$\pm 12.90 \pm 0.93$ mm	± 12 mm
VERT	-3.23 ± 0.78 mm	-4 mm

TABLE IV
ESTIMATION ERRORS FOR THE DIFFERENT METHODS

Methods	$mean \pm std$
Affine	2.42 ± 0.84 mm
Demons	1.77 ± 0.65 mm
B-splines	1.72 ± 0.48 mm
AC-PC	1.96 ± 0.90 mm

TABLE V
AC–PC ATLAS-BASED ERROR DECOMPOSITION

Components	$mean \pm std$
AP	0.95 ± 0.85 mm
LAT	1.02 ± 0.78 mm
VERT	0.86 ± 0.66 mm
AP-LAT	1.55 ± 0.96 mm

B. Atlas-Based Targeting and Registration Algorithms Versus Experts

The statistics, mean, and unbiased standard deviation, of the errors committed when applying the four methods to locate the STN, as described in Section II-D, are shown in Table IV. For the case of AC-PC atlas-based method, we have decomposed and computed the error in its three coordinates referred to the MCP point as well as the error considering only the laterality and the anteroposterior variability (see Table V) given that during the surgery a vertical correction of the electrode's location can partially be performed following the introduction trajectory path.

In order to compare the results, a one-way analysis of variance (ANOVA) test of the hypothesis that the errors came from distributions with equal means was performed over the errors produced by each method and by the experts at a 5% significance level. ANOVA determines the amount of variability in groups of data and shows whether the variability between groups is greater than within groups. This means that ANOVA provides a compact measure of the intergroup variance divided by the intragroup variance. In Fig. 5(a), a statistical box plot produced by this test is shown, as well as the result of a multicomparison test of the means [see Fig. 5(b)]. Two key conclusions can be drawn from these results. First, the mean errors committed with B-splines, demons and AC–PC based methods and the mean errors committed by the experts (expert variability) are not significantly different. Second, the mean error committed with affine registration is significantly different from the B-splines method and from the experts. Given the visual results of Fig. 5(a) and the statistical tests of the equality of means showed in Fig. 5(b), an F-test at a 5% significance level of the hypothesis that the set of errors generated by each method and by the experts come from distributions with equal variance was performed. The results, hypothesis, p-values and confidence intervals are shown in Table VI. The main conclusion is that the B-splines method produces an estimation error

(a) ANOVA statistical box plot
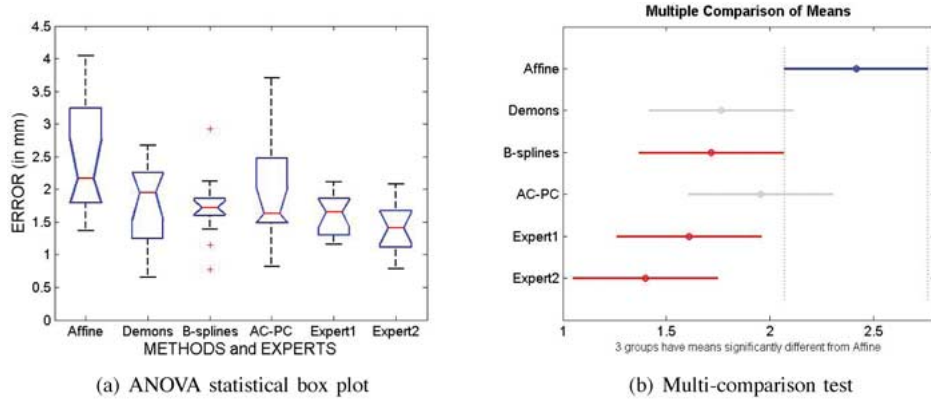
(b) Multi-comparison test

Fig. 5.   Statistical tests of the errors committed using different methods and by the experts (using the *anova1* and *multcompare* functions of MATLAB). (a) ANOVA statistical box plot. (b) Multicomparison test.

TABLE VI
F-Test of the Hypothesis of Equal Variance. Atlas-Based and Automatic Registration Methods Versus Experts. R Means That the Hypothesis was Rejected

F-test (5%)	H	p-value	CI
Affine vs Demons	–	0.367	[0.53 5.19]
Affine vs B-splines	–	0.059	[0.95 9.27]
Affine vs AC-PC	–	0.788	[0.27 2.67]
Affine vs Expert1	R	0.00051	[2.67 25.94]
Affine vs Expert2	R	0.0073	[1.57 15.25]
Demons vs B-splines	–	0.309	[0.57 5.55]
Demons vs AC-PC	–	0.244	[0.16 16.0]
Demons vs Expert1	R	0.0067	[1.60 15.54]
Demons vs Expert2	–	0.0627	[0.94 9.14]
AC-PC vs B-splines	R	0.0032	[1.11 10.79]
AC-PC vs Expert1	R	0.00022	[3.11 30.19]
AC-PC vs Expert2	R	0.0036	[1.83 17.76]
B-splines vs Expert1	–	0.074	[0.89 8.71]
B-splines vs Expert2	–	0.381	[0.52 5.12]
Expert1 vs Expert2	–	0.350	[0.18 1.83]

of the STN location that is statistically not different in mean and in variance to the experts variability.

In Fig. 6, we show the projection of each STN estimation (in red) onto the reference subject (in black) using the B-splines registration algorithm. Each point is represented by a circle of 1 mm of radius whose coordinates are projected onto the three orthogonal planes passing through the reference subject point coordinates (in black) in order to visualize the points in each view. The estimated targets are located very close to the *real target* and forming tight clouds of points showing that this kind of automatic estimation is reliable and well suited for this application.

C. Segmentation-Based Registration Versus Experts

The same statistics as above were computed for the segmentation-based method. In this case, we did not use a unique STN reference per side. A leave-one-out procedure was used where each patient is used once as reference or atlas. These statistics can be seen in Table VII, where we show the mean and unbiased standard deviation of the estimation error committed when

registering the images with an affine registration and the same statistics for the case of using only a structure or a combination of them with a tandem of affine-demons nonrigid registration. The corresponding box plot generated with the ANOVA test at a 5% significance level and the resulting multiple comparison of means are shown in Fig. 7(a) and (b), respectively. We can see that three different combinations give a mean error statistically different from the affine registration alone: L-T; L-C; L-T-C. From the results, we can conclude that the lateral ventricle has at least the same influence in the STN location as the third ventricle. This is a key conclusion given that usual atlas-based targeting procedures adjust the preoperative target using mainly the width of the third ventricle.

D. Validation and Comparison of All the Methods

Finally, an all-together validation was performed in order to compare the experts ability versus atlas-based, segmentation-based and automatic registration algorithms. In this case, the same reference subjects as in Section II-D were used. The numerical results ordered by decreasing mean error are shown in Table VIII. By simple inspection of the numerical results, it is not trivial to obtain strong conclusions. Is there any significant difference between the L-T case and the L-T-C? Can we ignore the interpeduncular cistern? Are these combinations better than the B-splines method?, etc. An ANOVA test at a 5% significance level and a multiple comparison test of means was performed over the whole set of estimation errors [see Fig. 8(a) and (b)]. We can see from these statistical tests of all the methods at the same time that if we use the segmented lateral and third ventricles, we obtain a mean error that is statistically different from the affine registration method while it is not statistically different from the experts. The use of the interpeduncular cistern does not contribute significantly to the reduction of the mean estimation error. Moreover, the use of a small region of interest (ROI) dramatically reduced the computation time compared to whole-image registration algorithms. On the other hand, the performance of the whole-image automatic registration using B-splines is no longer significantly different from affine registration regarding the mean estimation error of the STN location. Nevertheless, an F-test at a 5% significance level of the hypothesis that the set of errors generated by each method and by the experts come from distributions
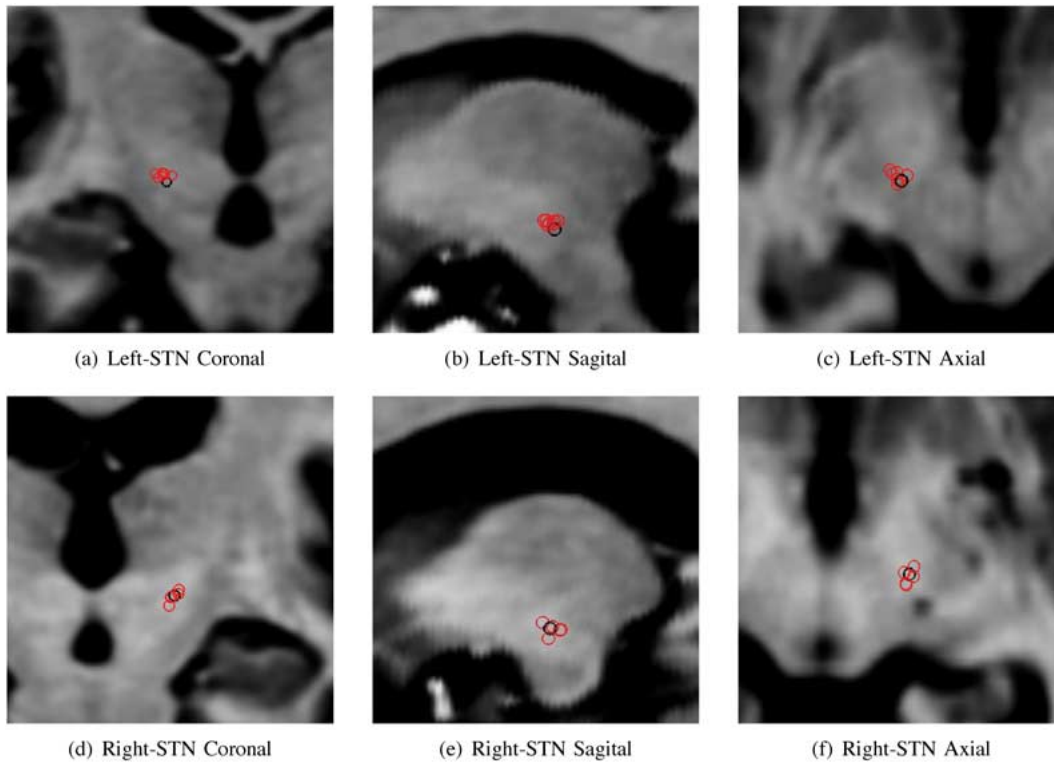
Fig. 6. STN estimation using B-splines free-form deformation algorithm (in red). Ground truth in black. (Each point represented by a circle of radius 1 mm). (a) Left-STN coronal. (b) Left-STN sagital. (c) Left-STN axial. (d) Right-STN coronal. (e) Right-STN sagital. (f) Right-STN axial.

TABLE VII
ESTIMATION ERRORS USING DIFFERENT SEGMENTED FEATURE-STRUCTURES

Structures	$mean \pm std$
Affine	2.31 ± 1.02 mm
L	1.97 ± 0.90 mm
T	2.01 ± 0.88 mm
C	2.12 ± 0.94 mm
T-C	1.97 ± 0.86 mm
L-T	1.82 ± 0.83 mm
L-C	1.88 ± 0.85 mm
L-T-C	1.76 ± 0.82 mm

TABLE VIII
STATISTICS OF THE ESTIMATION ERRORS COMMITTED WITH WHOLE-BRAIN
REGISTRATION, ATLAS-BASED METHOD, SEGMENTATION-BASED
REGISTRATION, AND INTER-EXPERT VARIABILITY

Methods, Features and Experts	$mean \pm std$
Affine	2.42 ± 0.84 mm
C	2.00 ± 0.72 mm
AC-PC	1.96 ± 0.90 mm
T	1.80 ± 0.69 mm
Demons	1.77 ± 0.65 mm
T-C	1.74 ± 0.71 mm
B-splines	1.72 ± 0.48 mm
L	1.70 ± 0.80 mm
L-C	1.67 ± 0.72 mm
Expert1	1.61 ± 0.29 mm
L-T	1.58 ± 0.71 mm
L-T-C	1.55 ± 0.73 mm
Expert2	1.40 ± 0.38 mm

with equal variance have shown that all the segmentation-based combinations displayed statistical difference in variance from, at least, one of the expert's variability, as shown in Table IX (only tests rejecting the hypothesis are displayed). In Fig. 9, we show the coronal projection of each estimated STN (in red) onto the reference subject (in black) using a L-T segmentation-based method. From this, we can see that this method provides an accurate estimation of the STN target.

IV. DISCUSSION AND CONCLUSION

The validation study presented here considered a critical clinical topic. Validation issues are crucial especially in DBS procedures where targets are small and their localization difficult. An accurate preoperative targeting can greatly reduce the operating time by decreasing the necessary corrections and adjustments required to attain a successful surgery. We discussed the intrarater and interrater variability in targeting clearly visible STNs. Then, an automated targeting technique was proposed that allows both the comparison and validation of different registration methods and to estimate their suitability for DBS targeting purposes. We have also studied the influence of neighboring structures in the STN location through a registration process. Some of the methods tested compared favorably to the expert's ability while others revealed that they were not suited for this application. One of the main conclusions we can drawn from this work is that automatic STN localization is possible and as accurate as the methods currently used.
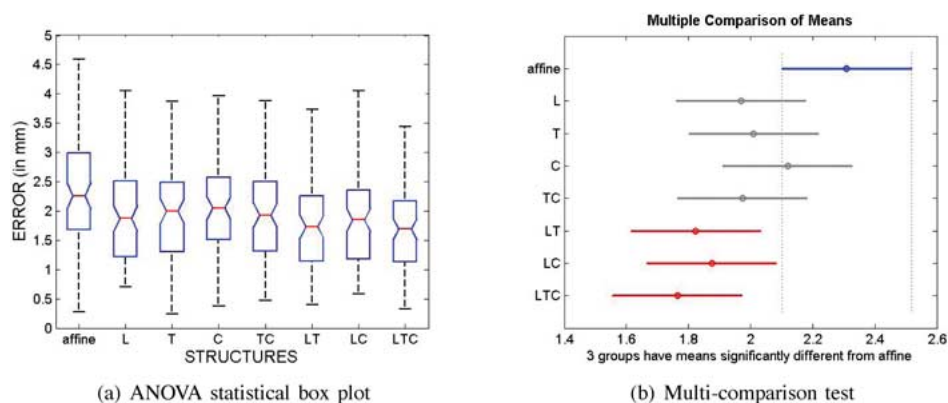
Fig. 7. Statistical tests of the errors committed using different segmented structures (using the *anova1* and *multcompare* functions of MATLAB). (a) ANOVA statistical box plot. (b) Multicomparison test.
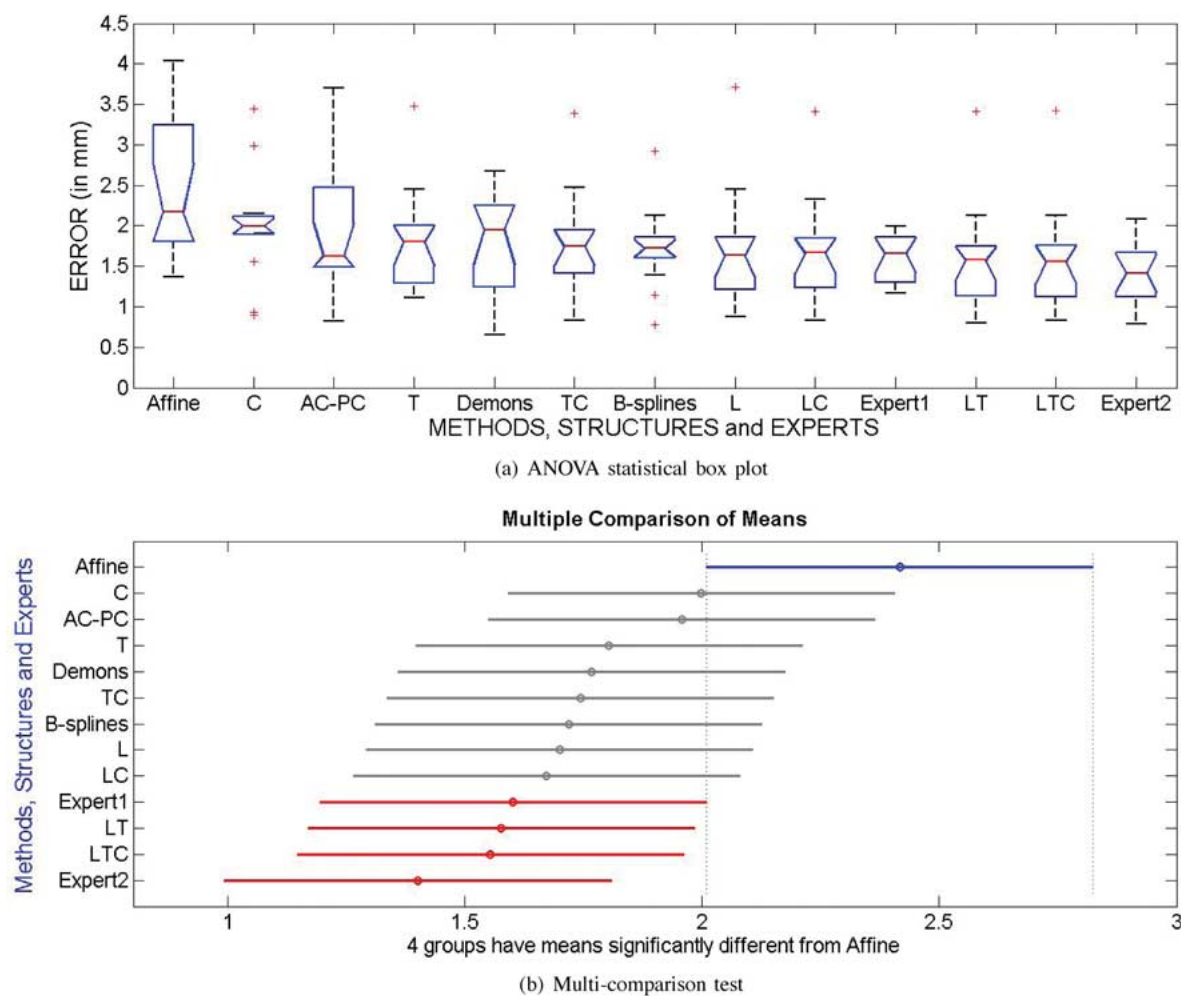

Fig. 8. Statistical tests of the errors committed using different atlas-based, segmentation-based, and automatic registration algorithms versus expert's variability (using the *anova1* and *multcompare* functions of MATLAB). (a) ANOVA statistical box plot. (b) Multicomparison test.

Among the AC–PC atlas-based and automatic nonrigid registration methods applied to the whole brain images, the B-splines algorithm demonstrated an extremely good performance with the smallest mean error and unbiased standard deviation, closely followed by demons and AC–PC atlas-based methods. Visually, numerically and statistically, the B-splines method produces estimations of the target which are comparable to the experts' variability. The estimated targets project to tight clusters showing the robustness of this kind of estimation method. However, statistical tests showed that global affine registration is not accurate enough for our application. Although the AC–PC atlas-based method shows an acceptable performance, it does not take into account the interpatient variability, which may be critical at a single patient level for an accurate targeting. We
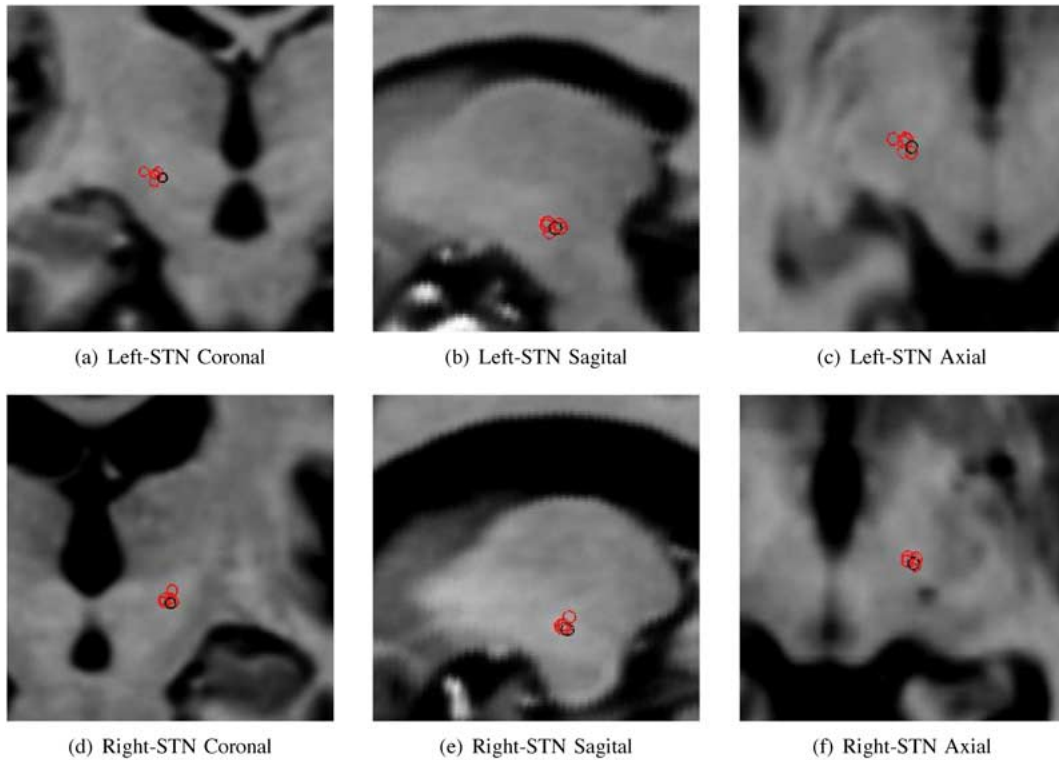
Fig. 9. STN estimation using a tandem of affine-demons registration using the segmented lateral and third ventricles. (Each point represented by a circle of radius 1 mm). (a) Left-STN coronal. (b) Left-STN sagital. (c) Left-STN axial. (d) Right-STN coronal. (e) Right-STN sagital. (f) Right-STN axial.

TABLE IX
F-TEST OF THE HYPOTHESIS OF EQUAL VARIANCE. SEGMENTATION-BASED METHOD VERSUS EXPERTS

F-test (5%)	H	p-value	CI
L vs Expert1	R	0.0018	[2.10 20.35]
L vs Expert2	R	0.0214	[1.23 11.97]
T vs Expert1	R	0.0081	[1.53 14.90]
C vs Expert1	R	0.0057	[1.66 16.12]
T-C vs Expert1	R	0.0065	[1.61 15.66]
L-T vs Expert1	R	0.0058	[1.65 16.02]
L-C vs Expert1	R	0.0053	[1.68 16.34]
L-T-C vs Expert1	R	0.0047	[1.73 16.76]
L-T-C vs Expert2	R	0.0471	[1.01 9.86]

have seen that the variance of the estimation error generated by this method is significantly different from the variance of the B-splines method and from the variance of the experts.

A study of surrounding visible easy-segmentable structures has revealed that the corresponding lateral ventricle is at least as influential as the third ventricle referred to the STN location while the use of the interpeduncular cistern does not contribute significatively to the improvement of the STN estimation compared to the former two. This is a crucial conclusion given that atlas-based targeting is usually modulated by the width of the third ventricle.

The use of an affine registration of the brain images followed by a segmentation-based demons registration of the binary mask of specific structures demonstrated an improvement of the results given by the whole brain image nonrigid registration methods. The use of a binary mask of the lateral and third ventricles provides a mean estimation error that is significantly different from the error given by affine registration but comparable to the expert's ability. Moreover, this method significantly reduces the computation time given that the non-rigid registration is computed on a small region of interest containing the binary mask and that the demons points to be considered are only those placed on the surface of the mask. It can also be fully automatic by using *a priori* knowledge of shape in the segmentation step.

The fact that the subthalamic area lacks contrast in MR T1-weighted images provides, through the validation scheme, a useful tool to study the propagation properties of the algorithms and methods under study.

Although the choice of the STN of reference can influence the results, the atlas-based and the whole-brain nonrigid registration methods were tested using six different combinations of STN-pairs (left and right references) and the results, numerically and statistically, were similar. The two subjects used as a reference for the right and left sides present large anatomical differences but the results obtained with the nonrigid registration algorithms did not show any significant differences regarding the side of the brain considered, demonstrating the strength of these methods to recover inter-patient anatomical variability.

However, some weak points of this study must to be noted. The study was performed using clearly visible STNs on the T2-weighted images. The rarity of this kind of images implies that a small sample group was considered (16 STNs). Although only MR T2-weighted slices clearly crossing the STN were considered, the spatial resolution of the T2 images introduces some errors in the definition of the real center of the STN. The spatial distortions present in MR images may introduce some

errors spreading the clusters of the estimated STNs. Reporting the real targets from the T2 images to the T1 space through rigid registration can also introduce some errors even if the images are acquired with the stereotactic frame fixed to the patient's skull which helps in driving the transformation. On the other hand, the AC–PC atlas-based method needs manual AC–PC point identification which introduces some variability error. All these possible sources of errors will be further investigated. Nevertheless, these limitations are also applicable to the work performed by other research groups [15]–[18], where most of these weaknesses are also present. They construct the ground-truth indirectly by registering different kinds of atlases: anatomical, electrophysiological, and/or histological and using final electrode's coordinates, which may introduce also additional errors. The registration of an anatomical segmented atlas to a patient's MR image has to consider the inherent errors of any registration process and the limitations of anatomical atlases (poor generalization, lack of morphometric information, spatial inhomogeneity, etc). The use of a postmortem brain needs volumetric reconstruction of the sliced brain and to solve a complex postmortem-MR registration problem, due to deformations produced by the process of extracting the brain from the cadaver, and by the slicing process itself. Electrophysiological recordings are useful and necessary for identifying the subthalamic region but different locations along the same track can produce similar responses of difficult interpretation. The final coordinates of the electrodes indicate that this location produces a good response to stimulation but we have to make the assumptions that the surgical team is able to place the electrode within the STN and that the intraoperative guidance system provides an accurate position of the electrodes. The main limitations of our work is the lack of a significantly large patient population with clearly visible STNs on T2-weighted series and a scarcity of neurosurgeons and radiologists with a high level of expertise and a wide experience in DBS targeting.

Future work will focus on including more patients, experts and registration methods, mainly local ones. The proposed automatic estimation of the STN can be used as a first and fast preoperative targeting procedure that can be refined by the neurosurgeon criterion and intraoperatively by electrophysiological recordings and macrostimulation. It can also be a useful and powerful tool for institutions without the expertise of leading research sites.

REFERENCES

[1] J. West, J. M. Fitzpatrick, M. Y. Wang, and B. M. Dawant *et al.*, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *Proc. SPIE Med. Imag. 1996*, vol. 2710, 1996.

[2] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, 1998.

[3] P. A. van den Elsen, E. J. D. Pol, and M. A. Viergever, "Medical image matching—A review with classification," *IEEE Eng. Med. Biol.*, vol. 12, no. 1, pp. 26–39, Jan. 1993.

[4] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[5] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, Jul. 2004.

[6] P. Hellier, C. Barillot, I. Courouge, B. Gibaud, G. Le Goualher, D. L. Collins, A. Evans, G. Malandain, N. Ayache, G. E. Christensen, and H. J. Johnson, "Retrospective evaluation of intersubject brain registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1120–1130, Sep. 2003.

[7] D. Collins and A. Evans, "ANIMAL: Validation and applications of non-linear registration-based segmentation," *Int. J. Pattern Recogn. Artif. Intell.*, vol. 11, pp. 1271–1294, Dec. 1997.

[8] A. L. Benabid, P. Pollak, A. Louveau, S. Henry, and J. de Rougemont, "Combined (thalamotomy and stimulation) stereotactic surgery of the VIM thalamic nucleus for bilateral Parkinson disease," *Appl. Neurophysiol.*, vol. 50, no. 1–6, pp. 344–346, 1987.

[9] C. Hamani, J. A. Saint-Cyr, J. Fraser, M. Kaplitt, and A. M. Lozano, "The subthalamic nucleus in the context of movement disorders," *Brain*, vol. 127, pp. 4–20, 2004.

[10] P. Limousin, P. Pollak, A. Benazzouz, D. Hoffmann, E. Broussolle, J. E. Perret, and A. L. Benabid, "Bilateral subthalamic nucleus stimulation for severe Parkinson's disease," *Movement Disorders*, vol. 10, pp. 672–674, 1995.

[11] A. L. Benabid, A. Koudsie, A. Benazzouz, V. Fraix, A. Ashraf, J. F. Le Bas, S. Chabardes, and P. Pollak, "Subthalamic stimulation for Parkinson's disease," *Arch. Med. Res.*, vol. 31, pp. 282–289, 2000.

[12] G. Schaltenbrand and W. Wahren, *Atlas for Stereotaxy of the Human Brain*, 2nd ed. New York: Thieme, 1977.

[13] J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme, 1988.

[14] C. Pollo, R. Meuli, P. Maeder, F. Vingerhoets, J. Ghika, and J. G. Villemure, "Subthalamic nucleus deep brain stimulation for parkinson's disease: Magnetic resonance imaging targeting using visible anatomical landmarks," in *Proc. Meeting Amer. Soc. Stereotactic Functional Surg.*, 2003, vol. 80, pp. 76–81.

[15] E. Bardinet, D. Dormont, G. Malandain, M. Bhattacharjee, B. Pidoux, C. Saleh, P. Cornu, N. Ayache, Y. Agid, and J. Yelnik, "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 *Proc. MICCAI*, 2005, vol. 3750, pp. 385–393.

[16] M. M. Chakravarty, A. F. Sadikot, J. Germann, G. Bertrand, and D. L. Collins, "Anatomical and electrophysiological validation of an atlas for neurosurgical planning," in *Proc. MICCAI*, 2005, vol. 3750, LNCS, pp. 394–401.

[17] T. Guo, K. W. Finnis, A. G. Parrent, and T. M. Peters, "Development and application of functional databases for planning deep-brain neurosurgical procedures," in *Proc. MICCAI*, 2005, vol. 3749, pp. 835–842.

[18] P. F. D'Haese, E. Cetinkaya, C. Kao, J. M. Fitzpatrick, P. E. Konrad, and B. M. Dawant, "Toward the creation of an electrophysiological atlas for the pre-operative planning and intra-operative guidance of deep brain stimulators (DBS) implantation," in *Proc. MICCAI*, 2004, vol. 3216, pp. 729–736.

[19] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[20] M. Bach, T. Butz, E. Solanas, O. Cuisenaire, and J.-P. Thiran, MIP: "The ITS-EPFL medical image processing library", EPFL, ITS Tech. Rep. 02.08, Aug. 2002.

[21] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," in *Proc. 5th ICCV*, 1995, vol. 16, pp. 16–23.

[22] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," in *Information Processing in Medical Imaging*. Norwell, MA: Kluwer, 1995, pp. 263–274.

[23] M. Wall, "GAlib 2.4.6: A C++ library of genetic algorithm components" Massachusetts Inst. Technol., Cambridge [Online]. Available: http://lancet.mit.edu/ga/

[24] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*. London, U.K.: Cambridge Univ. Press.

[25] J.-P. Thirion, "Non-rigid matching using demons," in *Proc. IEEE CVPR*, Jun. 1996, pp. 245–251.

[26] ——, "Image matching as a diffusion process: An analogy with maxwell's demons," *Med. Image Anal.*, vol. 2, pp. 243–260, 1998.

[27] M. Bach, "Atlas-based segmentation and classification of magnetic resonance brain images," Ph.D. dissertation, Ecoles Polytechniques Federale de Lausanne, Lausanne, Switzerland, 2004.

[28] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[29] W. Gropp, E. Lusk, and A. Skjellum, *Using MPI: Portable Parallel Programming With the Message-Passing Interface*. Cambridge, MA: MIT Press, 1999.

[30] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[31] X. Bresson, P. Vandergheynst, and J.-P. Thiran, "A variational model for object segmentation using boundary information and shape prior driven by the Mumford-Shah functional," *Int. J. Comput. Vision*, vol. 28, pp. 145–162, 2006.