ORIGINAL ARTICLE

Validation of experts versus atlas-based and automatic registration methods for subthalamic nucleus targeting on MRI

**F. Javier Sanchez Castro · Claudio Pollo ·
Olivier Cuisenaire · Jean-Guy Villemure ·
Jean-Philippe Thiran**

Abstract *Objects* In functional stereotactic neurosurgery, one of the cornerstones upon which the success and the operating time depends is an accurate targeting. The subthalamic nucleus (STN) is the usual target involved when applying deep brain stimulation for Parkinson's disease (PD). Unfortunately, STN is usually not clearly visible in common medical imaging modalities, which justifies the use of atlas-based segmentation techniques to infer the STN location. *Materials and methods* Eight bilaterally implanted PD patients were included in this study. A three-dimensional T1-weighted sequence and inversion recovery T2-weighted coronal slices were acquired pre-operatively. We propose a methodology for the construction of a ground truth of the STN location and a scheme that allows both, to perform a comparison between different non-rigid registration algorithms and to evaluate their usability to locate the STN automatically. *Results* The intra-expert variability in identifying the STN location is 1.06 ± 0.61 mm while the best non-rigid registration method gives an error of 1.80 ± 0.62 mm. On the other hand, statistical tests show that an affine registration with only 12 degrees of freedom is not enough for this application. *Conclusions* Using our validation–evaluation scheme, we demonstrate that automatic STN localization is possible and accurate with non-rigid registration algorithms.

F. J. Sanchez Castro (✉) · C. Pollo · O. Cuisenaire · J.-P. Thiran
Signal Processing Institute,
École Polytechnique Fédérale de Lausanne (EPFL),
1015 Lausanne, Switzerland
E-mail: francisco.sanchezcastro@epfl.ch

C. Pollo · J.-G. Villemure
Department of Neurosurgery,
Centre Hospitalier Universitaire Vaudois (CHUV),
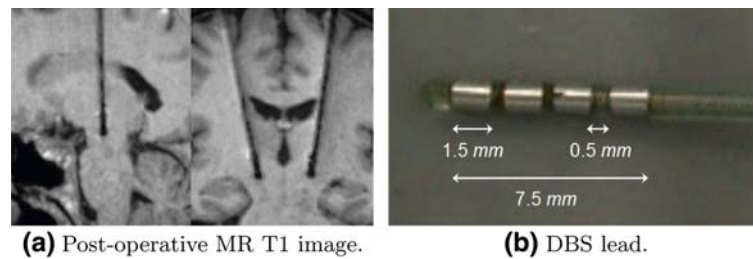Lausanne, Switzerland

Introduction

From the late 1960s, with the introduction of *levodopa*, medications have dominated the treatment of Parkinson's disease (PD). Unluckily, medical treatment has significant shortcomings. Benefits go down with time. Many patients suffer from *wearing-off*, a decrease of the medical effect following each dose. After some years patients often become less tolerant to the drug. Many of them develop drug-induced effects, like severe rigidity and tremor (*off* periods) and severe *dyskinesias*, involuntary twisting and writhing movements (*on* periods). These problems with medical therapy and the increase of basal ganglia circuitry knowledge have led to the development of surgical-based techniques. We can distinguish two types of approaches: lesion-based techniques and deep brain stimulation (DBS). The first ones consist in the surgical destruction of a small amount of brain tissue in some regions (specific parts of some basal ganglia). Amongst the lesion-based, thalamotomy and pallidotomy have shown good performance in treating tremor in one side of the body (the opposite side to the lesion side) and reducing dyskinesias, but they cannot be applied bilaterally because bilateral lesion may produce severe undesirable and permanent effects on speech, behavior or cognition. DBS can be performed safely on both sides of the brain and, contrary to lesion-based techniques, DBS 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. After hundreds of surgical interventions the subthalamic nucleus (STN) has turned out to be the most effective target for DBS.

A typical DBS procedure starts with the placement of the stereotactic head frame, fixed to the patient's skull, that will be used as a coordinate reference. Next an imaging study is done in order to select the target to stimulate and to plan

Fig. 1 Post-operative validation
problems

(a) Post-operative MR T1 image.   **(b)** DBS lead.

the trajectories for introducing the electrodes. Usually two kinds of images are taken to be able to visualize different tissues, MR T1-weighted and MR T2-weighted images. In our state-of-the-art protocol, the selection of the STN target is performed on a coronal T2-weighted image acquired perpendicularly to the anterior commisune–posterior commisune (AC–PC) axis and crossing the anterior limit of the red nucleus. STN target selection depends on each institution. Common methods are the use of stereotactic atlases of the brain [1,2] and the use of visible surrounding anatomical landmarks [3]. Then the target coordinates are reported to the T1-weighted image where trajectories are planned. Once, in the operating room, the head frame is fixed to the operating table, a small hole is drilled into the patient's skull according to pre-operative trajectories. Because of the difficulty to see the STN directly from medical images, the pre-operative target is only an estimation of the real location. The electrode's location has to be adjusted intra-operatively using electrophysiological recordings and from stimulation tests.

As we dispose of the post-operative images for all the patients involved in this study, it is easy to imagine a straightforward way to assess STN location: the final coordinates of the electrodes. Unfortunately, post-operative MR images do not provide accurate localization of the electrode's tip due to the artifacts generated in the images by the electrode itself, as can be seen in Fig. 1a. The DBS lead is composed of an insulated wire with four platinum iridium contact electrodes on its tip, each contact measuring 1.5 mm in length separated by 1.5 or 0.5 mm (see Fig. 1b). Ideally, one contact should be located within the STN or two contacts should be surrounding it. It is very difficult to assert that certain contact is within the target or in the closest neighborhood. The use of micro-electrode recordings allows to guess that we are more or less within the STN, but it is a very small structure. Therefore, these possibilities do not provide enough accuracy.

The goal of the study reported in this paper is to voice the possibility to automatically locate the STN using existing registration techniques and to compare their performance and usability. Knowing the location of the STN for all our patients' database, a patient is chosen as atlas and each individual patient mapped back to the atlas using the registration methods considered in this study to obtain an estimation of

the STN, which is compared to the real location given by the expert.

Material and methods

Data

For each patient of our 34 bilaterally implanted Parkinsonian patients' database, two kinds of images were taken pre-operatively: a 3D T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) MRI sequence (Siemens Vision®, 1.5 T; Erlangen, Germany) TR 9.7 ms, TE 4 ms, number of slices/slice thickness: 164/1.40 mm, FOV 280×280, matrix 256×256, pixel size 1.09×1.09 mm^2, and few coronal slices (due to the time taken for this kind of imaging sequence) of an inversion recovery (IR) T2-weighted, TR 2,560 ms, TE 4 ms, number of slices/slice thickness: 7/3 mm, FOV 300×300, matrix 512×512, pixel size 0.59×0.59 mm. Taking profit from the fact that in some rare cases the STN is visible in MR T2-weighted images, a reference can be constructed and used as a ground truth. To do this neurosurgeon experts have been asked to select patients with clearly visible STN in MR T2-weighted images amongst our patients' database. After exhaustive inspection eight patients were selected (16 STNs).

Reference construction and validation scheme

An experimented neurosurgeon with wide experience in PD surgery and targeting has been asked to click the target point (two STN) for each selected T2 series. This process has been repeated five times for each patient at different days to avoid that the expert be influenced by previous targeting choices. With this data we were able to compute a statistical mean target point coordinates which we call the *real targets* (see Fig. 2).

Amongst the eight selected subjects, the one with the most clearly visible STN has been selected as a reference subject, both for the right and left sides. Intra-expert variability statistics can also be extracted. Using these data we can consider our reference subjects as an atlas. Then, by non-rigidly registering the atlas with the patient image, we obtain an
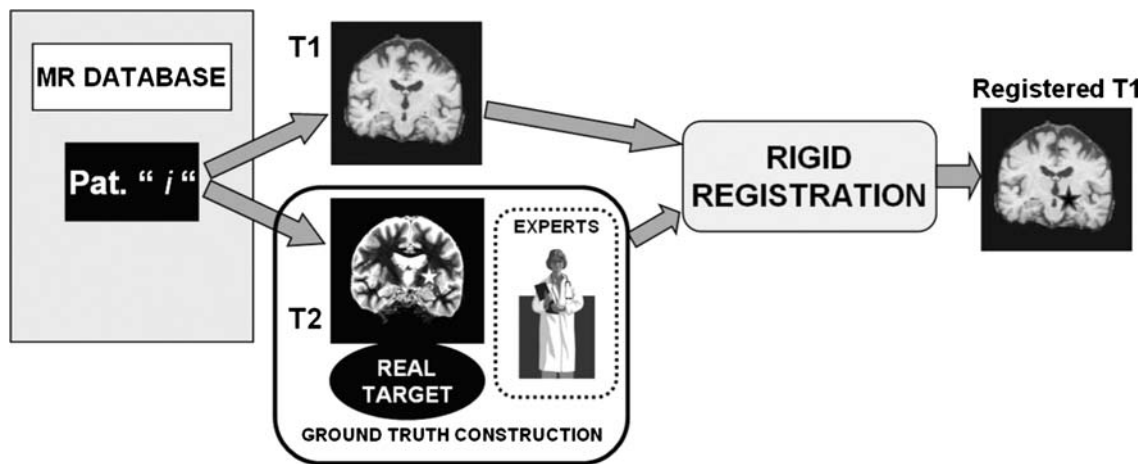
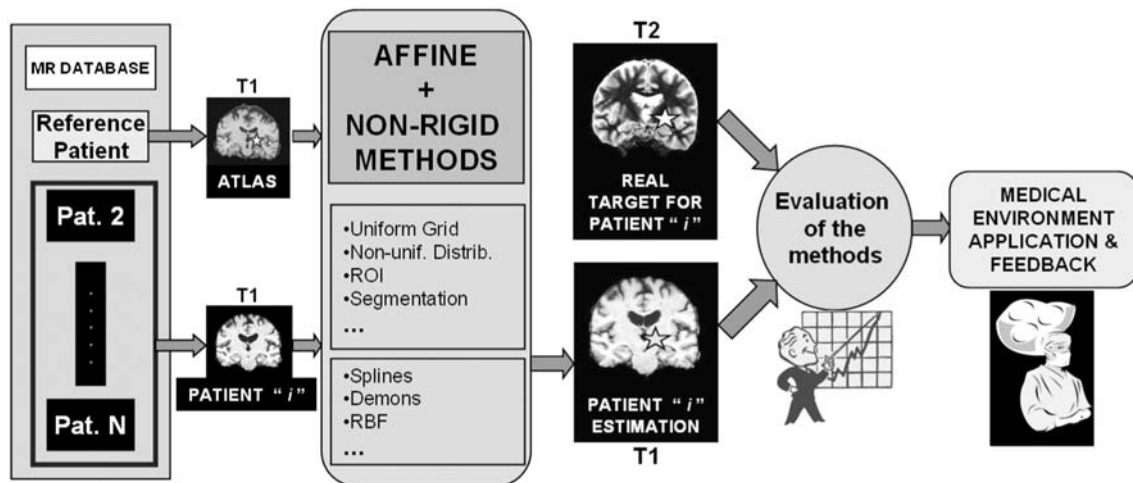**Fig. 2** Ground truth construction. Target localization

Fig. 3 Validation scheme. Target estimation

estimation of the STN position in the patient. The procedure is as follows. The first step is to report the MR T2 coordinates to MR T1 coordinates, for each patient, to obtain the location of the real targets in the T1 domain. To do this a rigid registration between the two modalities (six degrees of freedom: translation and rotation) has been performed [4] as shown in Fig. 2. The next step consists in applying each of the registration algorithms under study to get an estimation of the position of the STN. This estimation is obtained by registering the MR T1-weighted image of the atlas with the MR T1-weighted image of the patient under study. In this work four methods were considered (see AC–PC atlas-based targeting and automatic registration algorithms). Finally, repeating this for the seven datasets (14 STNs) involved (leaving out the references), Euclidean distances from estimated to real targets (seen as the mean target point given by the expert) for each STN are calculated and statistics are extracted to evaluate the performance of the different methods. This

evaluation–validation procedure is shown as a block diagram in Fig. 3.

AC–PC atlas-based targeting and automatic registration algorithms

To face up to image registration the following methods have been tested:

Atlas-based (AC–PC) targeting AC–PC referential together with brain atlases is one of the methods used to target the STN in medical environment, mainly when STN is not clearly visible in MR T2-weighted images which is the usual case. Neurosurgeons have to locate the anterior and posterior commissures (AC–PC points). Then, using a stereotaxic atlas and taking the midcommissural point (MCP) as the origin, we can estimate that the STNs are located at coordinates (following Schaltenbrand–Warren [1]): anteroposterior (AP) −3 mm,

Table 1 Expert targeting variability. All the STNs and references

Mean ± SD	All STNs (mm)	Left STN reference (mm)	Right STN reference
Paired	1.67 ± 0.98	1.64 ± 0.88	1.18 ± 0.52
Centroid	1.06 ± 0.61	1.10 ± 0.32	0.75 ± 0.38
MCP Distance	13.47 ± 1.37	12.38 ± 0.61	13.63 ± 0.54

lateral (LAT) ±12 mm (left side and right side) and vertical (VERT) −4 mm.

Affine registration We used an independent implementation based on the work of Maes et al. [4]. The 12 degrees of freedom (translation, rotation, scaling and shearing) are optimized in order to maximize the mutual information between the images to be registered [5,6] using a multiscale approach and a two-step optimization. First a global search using genetic algorithms [7] and next a local optimization using the steepest descent method [8]. Affine registration is also used as a pre-alignment step for non-rigid transformations described below.

Demons algorithm It is an intensity-based algorithm proposed by Thirion [9,10] and 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 semipermeable 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. 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 images of F and G, respectively. The deformation algorithm is applied by iterating in a hierarchical coarse-to-fine multiscale way. The smoothness of the displacement field is imposed by smoothing with a Gaussian filter of standard deviation σ (*elasticity parameter*) chosen empirically [11]. In our case, parameter σ has been chosen by exhaustive search, between 0.6 and 2.0 mm by steps of 0.2 mm, minimizing the distance between the estimated STNs and the real targets. Finally, σ of 1 mm has been chosen.

B-splines algorithm It is a mutual information-based free-form deformation algorithm whose displacement field is modeled as a linear combination of B-splines lying in a regular grid (uniformly spaced control points) similar to the method proposed by Rueckert et al. [12]. 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:

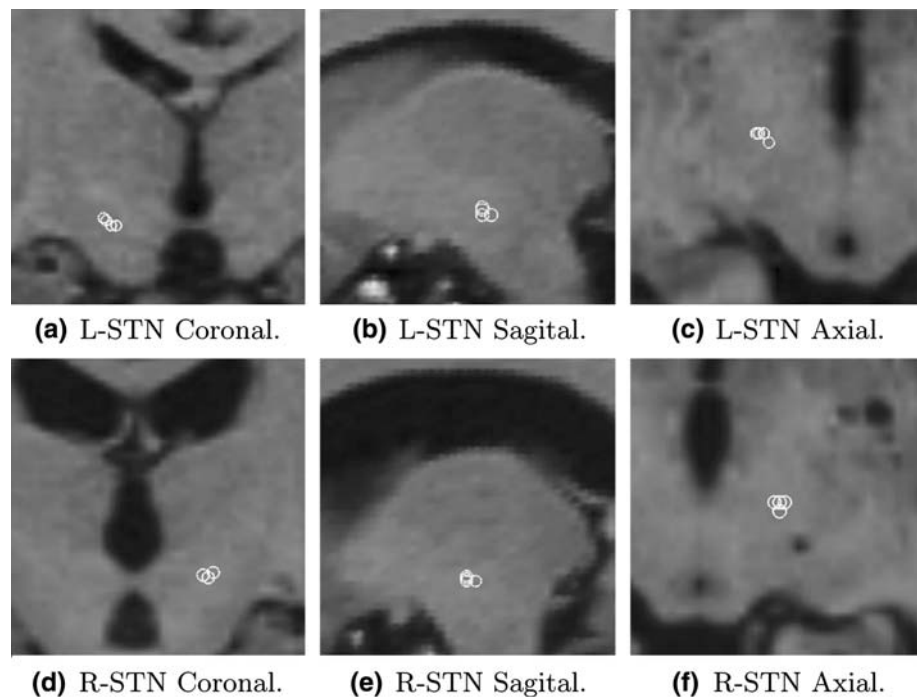$$d(x) = \sum_{k \in \mathbb{Z}} c(k)\beta^3(x - k),$$

where $c(k)$ are the B-spline coefficients, $\beta^3(*) = \beta^0 \times \beta^0 \times \beta^0 \times \beta^0(*)$, and β^0 a rectangular pulse. To speed up the optimization process, the algorithm has been implemented using the communication utilities for distributed memory architectures using the MPICH implementation of the message passing interface (MPI) [13]. The good interpolation properties and the suitability for multiscale processing of the B-splines are well known [14], and its deformability can be controlled by changing the spacing between the control points of the grid which we have set at 12 mm.

Results

Target selection and surgeon variability

Two main statistics can be extracted from the data given by the expert: intra-surgeon variability and the mean target point coordinates for each STN which will be used as a *real* target coordinates.

In order to evaluate the *repeatability* of the neurosurgeon targeting, we have proceeded in different manners. First we have computed, for each STN, the Euclidean distance between each couple of neurosurgeon's clicks (5 clicks per STN, 10 distances per STN, 160 distances for the 8 selected patients from the database) and we obtain the mean and standard deviation (SD) (using the unbiased estimator) of the so-called *paired variability*. The same information can be extracted computing the centroid for each STN cloud of points and calculating its Euclidean distances to each of these points. This gives the mean and unbiased SD of the so-called *centroid variability*. We also compute the mean distances from the targeted points to the MCP. These quantities allow us to get an idea of the surgeon variability and its accuracy when clicking over the pre-operative target and can be seen in the second column of Table 1. If we only consider the STNs used as a reference we obtain the paired and centroid variabilities as well as the mean distance to the MCP for the left and right sides showed in the third and fourth columns, respectively, of Table 1.

Fig. 4 Reference STN expert targeting

(a) L-STN Coronal.   **(b)** L-STN Sagital.   **(c)** L-STN Axial.

(d) R-STN Coronal.   **(e)** R-STN Sagital.   **(f)** R-STN Axial.

Table 2 Expert variability. Distance to MCP

Target	Mean ± SD (mm)	Target	Mean ± SD (mm)
STN 1	14.92 ± 1.52	STN 9	12.79 ± 0.84
STN 2	13.66 ± 1.69	STN 10	14.01 ± 0.29
STN 3	13.25 ± 0.26	STN 11	12.38 ± 0.61
STN 4	14.35 ± 0.34	STN 12	11.89 ± 0.91
STN 5	14.69 ± 0.68	STN 13	14.31 ± 1.67
STN 6	13.63 ± 0.54	STN 14	13.88 ± 0.79
STN 7	11.72 ± 0.84	STN 15	14.68 ± 0.83
STN 8	11.32 ± 0.40	STN 16	14.02 ± 0.46

Table 3 Mean STN coordinates referred to the MCP

Coordinates	Mean ± SD (mm)
AP	−3.29 ± 1.36
LAT	±12.57 ± 1.39
VERT	−3.16 ± 0.89

In Table 2, we show the expert variability referred to the distances to the MCP for each STN targeted, following the procedure described in Sect. Reference construction and validation scheme.

In Fig. 4, the points targeted by the expert for the left (Figs. 4a–c) and right (Figs. 4d–f) STNs chosen as a reference can be seen. For visualization purposes, each point is projected onto the three orthogonal planes passing through the centroid and showed using a circle (of radius 1 mm).

We can also decompose these distances in AP, LAT and VERT coordinates which allows comparing directly the mean STN location for our database with usual STN coordinates

from the atlases (e.g. [1,2]). In Table 3 we show the mean and SD of the coordinates referred to the MCP for the 16 STNs used in this study.

Target estimation and evaluation of the methods

The statistics (mean and unbiased SD) of the errors committed when applying these algorithms and methods to locate the STN, as described under AC–PC atlas-based targeting and automatic registration algorithms, are shown in Table 4. For the case of AC–PC atlas-based method we decompose and compute the error in its three coordinates or directions as well as the error considering only the LAT and the AP variability (see Table 5).

Table 4 Estimation errors

Methods	Mean ± SD (mm)
Affine	2.65 ± 1.25
Demons	1.97 ± 0.85
B-Splines	1.80 ± 0.62
AC–PC	1.98 ± 0.81

Table 5 Atlas-based errors' decomposition

Components	Mean ± SD (mm)
AP	0.98 ± 0.72
LAT	1.07 ± 0.65
VERT	0.92 ± 0.64
AP-LAT	1.57 ± 0.74

In order to compare the results, statistical tests have been done. In Fig. 5 a statistical box plot, generated by using an one-way analysis of variance, for the different methods tested is shown. A paired t-test of the hypothesis that the errors come from distributions with equal means has been performed over each pair of methods at a 5% significance level. The numerical results of the test (p-values and confidence intervals) are shown in Table 6. We obtain that the mean errors committed with B-splines, demons and AC–PC based methods are not significantly different. However, the results show that affine registration is significantly different from the other methods above.

Given the visual results of Fig. 5 and the numerical results produced by the t-test of the equality of means, we perform an F-test at a 5% significance level of the hypothesis that the set of errors generated by each method come from distributions with equal variance. The results are shown in Table 7. We can see that only the variance of the errors produced by the B-splines method is significantly different from the one produced by affine registration.

In Fig. 6, we show the projection of each STN estimation (in white) onto the reference subject (in red) using the B-
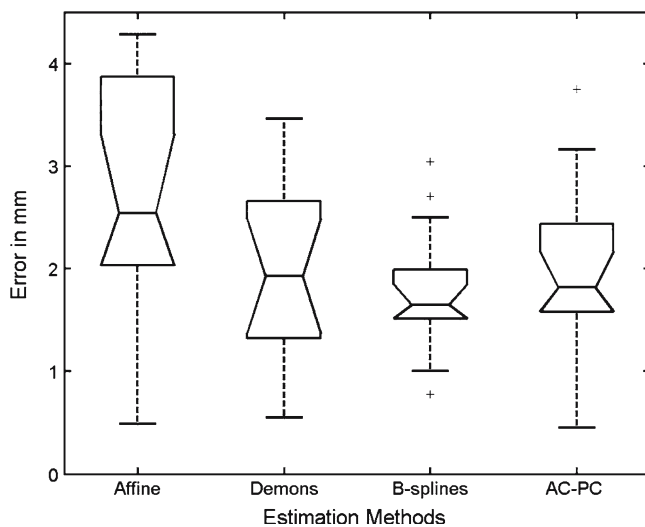

Fig. 5 Statistical box plot of the errors committed using different methods

Table 6 t-test of the hypothesis of equal means

t-test	Hypothesis	P-value	Confidence interval
Affine vs demons	Rejected	0.027	[0.090 1.272]
Affine vs B-splines	Rejected	0.047	[0.014 1.688]
Affine vs AC–PC	Rejected	0.033	[0.063 1.276]
Demons vs B-splines	–	0.454	[−0.304 0.643]
Demons vs AC–PC	–	0.964	[−0.569 0.546]
B-Splines vs AC–PC	–	0.563	[−0.841 0.479]

Table 7 F-test of the hypothesis of equal variance

F-test	Hypothesis	P-value	Confidence interval
Affine vs demons	–	0.173	[0.700 6.792]
Affine vs B-splines	Rejected	0.016	[1.317 12.784]
Affine vs AC–PC	–	0.130	[0.765 7.427]
Demons vs B-splines	–	0.267	[0.604 5.863]
Demons vs AC–PC	–	0.8744	[−2.849 −0.294]
B-splines vs AC–PC	–	0.340	[−5.362 −0.553]

splines registration algorithm. Each point is represented by a circle of 1 mm of radius whose coordinates are projected onto the reference subject point coordinates (in red) in order to visualize the points in each view and refer to the reference subject point. The estimated targets are located very close to the *real target* and form tight clouds of points showing that this kind of automatic estimation is reliable and well suited for this application.

Conclusions and perspectives

The main conclusion one can extract from the results we have obtained from this evaluation–validation work is that automatic STN location is possible and accurate. As we can see, by simple inspection of the numerical and visual results, the B-splines method shows the best performance with the smallest mean error and adjusted (unbiased) SD but is closely followed by demons and AC–PC methods. The points project on tight clusters showing the robustness of this kind of estimation method. The statistical tests have shown that global affine registration is not enough for our application and that there are no significant statistical differences, neither mean nor variance, between the errors committed using the B-splines, demons or AC–PC referential-based techniques. Note that only the variance of the B-splines method is statistically different from the one of the affine registration. Although AC–PC referential-based method shows an acceptable performance, its estimation needs AC and PC point's identification by an expert and does not take into account the inter-patient variability, which is very important at a single patient level. The fact that AC–PC shows such a good behavior can be explained because the selected patients follow well the *average* brain defined by the atlas. But what about patients with abnormal anatomy or damaged brains? Although the choice of the STN of reference can influence the results, the methods have been tested using six different combinations of STN pairs (left and right references) and the results were similar. The automatic targeting of the STN can be also used as a first and fast pre-operative target estimation that can be refined by the neurosurgeon criterion. The work in progress includes a study of *inter*-expert variability that will

Fig. 6 STN estimation using B-splines

(a) L-STN Coronal.   (b) L-STN Sagital.   (c) L-STN Axial.

(d) R-STN Coronal.   (e) R-STN Sagital.   (f) R-STN Axial.

Source: 2272x1704x801
Resolution 100x100x150 μm
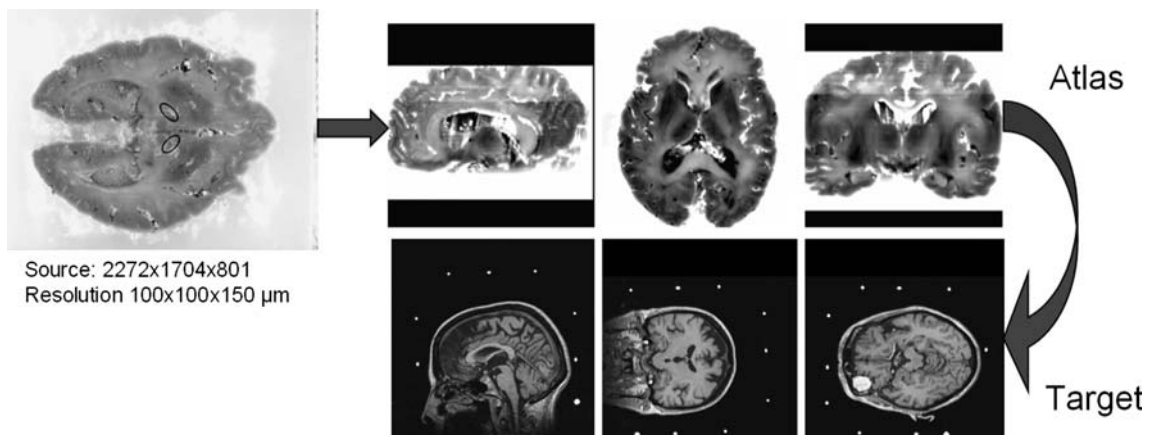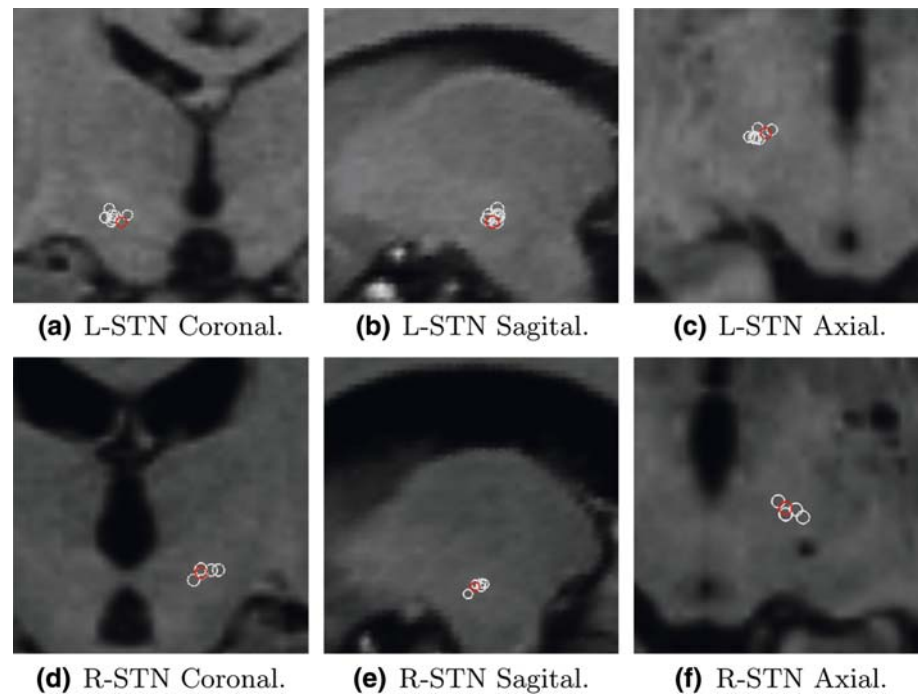
Atlas

Target

Fig. 7 Post-mortem atlas generation

provide a mean to obtain more accurate STN target atlases as well as a cross-comparison of experts and registration methods. As our patients' database increases and more experts as well as more registration techniques, mainly local ones, are included in the study, we will be able to establish stronger assessments and conclusions. An interesting and promising work we are developing is the construction of a post-mortem brain atlas (see Fig. 7). We dispose of a high resolution sliced dead Parkinsonian brain where the STN is well defined and easy to segment. We have also the pre- and post-operative MR T1- and T2-weighted studies of the same patient when he was alive. Using rigid registration we can reconstruct the 3D volume from the pictures of the dead brain and then project the segmented STN onto the MRI images by using accurate non-rigid registration algorithms. This will provide an MRI STN atlas that will be used as the reference in our proposed scheme. However, the process of bringing the dead brain and its corresponding alive MRI into enough spatial correlation is a very difficult task and still an open question.

References

1. Schaltenbrand G, Wahren W (1977) Atlas for stereotaxy of the human brain, 2 edn. Thieme, Stuttgart
2. Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Thieme, Stuttgart
3. Pollo C, Meuli R, Maeder P, Vingerhoets F, Ghika J, Villemure J-G (2003) Subthalamic nucleus deep brain stimulation for Parkinson's disease: magnetic resonance imaging targeting using visible anatomical landmarks. In: Proceedings of the meeting of the american society for stereotactic and functional surgery, New York, vol 80 No. 1–4
4. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P (1997) Multimodality image registration by maximization of mutual information. IEEE Trans Med Imag 16(2):187–198
5. Viola P (1995) Alignment by maximization of Mutual Information. M.I.T. Artificial Intelligence Technical Report, no. 1548
6. Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G (1995) Automated multi-modality image registration based on information theory. Information processing in medical imaging, Kluwer, Dordrecht, pp 263–274
7. Wall M GAlib 2.4.6: A C++ Library of genetic algorithm components. Massachusetts Institute of Technology (http://lancet.mit.edu/ga/)
8. Press W, Teukolsky S, Vetterling W, Flannery B Numerical recipes in C, the art of scientific computing. Cambridge University Press, London (http://www.nr.com)
9. Thirion J-P (1996) Non-rigid matching using demons. Computer vision and pattern recognition, 1996. In: Proceedings CVPR '96, IEEE Computer Society Conference, 18–20 June 1996, pp. 245–251
10. Thirion J-P (1998) Image matching as a diffusion process: an analogy with Maxwells demons. Med Imag Anal 2(3):243–260
11. Bach M (2003) Atlas-based segmentation and classification of magnetic resonance brain images. PhD thesis, EPFL, CH-1015 Thesis No. 2875
12. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ (1999) Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imag 18(8):712–721
13. Gropp W, Lusk E, Skjellum A (1999) Using MPI: portable parallel programming with the message-passing interface. MIT Press, Cambridge
14. Unser M (1999) Splines: a perfect fit for signal and image processing. IEEE Signal Process Mag 16:22–38