

Cross Validation of Experts Versus Registration Methods for Target Localization in Deep Brain Stimulation^{*}

F. Javier Sánchez Castro¹, Claudio Pollo^{1,2}, Reto Meuli³,
Philippe Maeder³, Meritxell Bach Cuadra¹, Olivier Cuisenaire¹,
Jean-Guy Villemure², and Jean-Philippe Thiran¹

¹ Signal Processing Institute, École Polytechnique Fédérale de Lausanne (EPFL),
CH-1015 Lausanne, Switzerland
`francisco.sanchezcastro@epfl.ch`

Departments of

² Neurosurgery and

³ Radiology, Centre Hospitalier Universitaire Vaudois (CHUV), Switzerland

Abstract. In the last five years, Deep Brain Stimulation (DBS) has become the most popular and effective surgical technique for the treatment of Parkinson's disease (PD). The Subthalamic Nucleus (STN) is the usual target involved when applying DBS. Unfortunately, the STN is in general not visible in common medical imaging modalities. Therefore, atlas-based segmentation is commonly considered to locate it in the images. In this paper, we propose a scheme that allows both, to perform a comparison between different registration algorithms and to evaluate their ability to locate the STN automatically. Using this scheme we can evaluate the expert variability against the error of the algorithms and we demonstrate that automatic STN location is possible and as accurate as the methods currently used.

1 Introduction

Deep Brain Stimulation (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 taken in order to select pre-operatively the target to be stimulated and to plan the trajectories for introducing the electrodes. Usually two kind 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 AC-PC axis and crossing the anterior limit of

^{*} This work is supported by the Swiss National Science Foundation under grant number 205320-101621.

the Red Nucleus. STN target selection depends on each institution. Common methods are the use of a stereotactic atlas of the brain [1] and the use of visible surrounding anatomical landmarks [2]. Then, the target coordinates are reported to a T1-weighted image where the 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 the pre-operative trajectories. Because of the difficulty to directly see the STN on medical images, the selected pre-operative target is only an estimation of the real location. Therefore the electrode's location has to be adjusted intra-operatively using electrophysiological recordings and from stimulation tests.

This paper focuses on three main goals. First of all, the construction of a reference which is used as a ground truth for the position of the STN. Secondly, the study of the *intra* and *inter*-expert variability in identifying the STN location visually. Thirdly, to evaluate the possibility of automatically locating the STN using existing registration techniques and to compare their performance and usability. The construction of a STN location ground truth from the experts knowledge is as follows. First, a patient is chosen as atlas and each individual patient mapped back to this atlas using various registration methods. Then, an estimation of the STN location is obtained and compared to the real location given by the ground truth. Using accurate registration algorithms we demonstrate that automatic STN localization is possible and as accurate as the methods currently used. As far as we know, only one study to evaluate a standard mutual information-based non-rigid registration algorithm for automatically locating the STN has been published [3]. As opposed to us, they use post-operative coordinates as a ground truth, making the assumptions that the surgical team is able to place the electrode within the STN and that the intra-operative guidance system provides the accurate position of the electrodes.

2 Materials and Methods

2.1 Data

For each patient of our bilaterally implanted parkinsonian patient's database (37 patients, 74 STNs) two kinds of images were acquired pre-operatively: a 3D 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 280x280, matrix 256x256, pixel size 1.09x1.09 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 2,560 ms, TE 4 ms, number of slices/slice thickness: 7/3 mm, FOV 300x300, matrix 512x512, pixel size 0.59x0.59 mm. Taking profit from the fact that in some rare cases the STN is visible in MR T2-weighted images, a reference is constructed and used as a ground truth. To do this, neurosurgeon experts were asked to select patients with clearly visible STN in MR T2-weighted images amongst our patient's database. After exhaustive inspection 8 patients were selected (16 STNs).

2.2 Ground Truth and Validation Scheme

Two experts, one neurosurgeon and one radiologist, both with wide experience in Parkinson’s disease (PD) surgery and targeting, have been asked to click the target point (2 STN per patient) for each selected T2 series. This process has been repeated 5 times for each patient at different days to avoid that the experts be influenced by previous targeting choices. With this data we were able to compute statistics about *intra* and *inter*-expert variability. *Inter*-expert targeting differences have turned out to not be statistically significant as shown in section 3.1. Thus, we can use the two sets of points to compute statistical mean target point coordinates, called *real targets*, and which are considered the ground truth. Amongst the 8 selected subjects, the experts have selected the one with the most clearly visible STN as a reference subject, both for the right and left sides. These data together with the real targets allow us to consider our reference subjects as an atlas. Then, by non-rigidly registering the 7 other patient’s images with the atlas, we obtain a projection of the STN of the patient in the atlas, an estimation of the STN position given by the registration method used. The procedure is as follows. First, for each patient, we perform a T1-T2 rigid registration [4] and project the real target coordinates into the T1 space. Next, we apply each of the 4 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 patient under study with the MR T1-weighted image of the atlas. Finally, repeating this for the 7 datasets (14 STNs) involved in this study (leaving out the reference subjects), Euclidean distances from estimated to real targets for each STN are calculated and statistics are extracted to evaluate the performance of the different methods.

2.3 Atlas-Based Targeting and Registration Algorithms

In this study, the following registration methods have been applied and tested because of their wide use in medical image processing:

- *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 as the origin the midcommissural point (MCP), this procedure estimates that the STNs are located at coordinates (following Schaltenbrand-Wahren [1]): anteroposterior (AP) $-3mm$, lateral (LAT) $\pm 12mm$ (left and right side) and vertical (VERT) $-4mm$.
- *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] using a multiscale approach and a two-step optimization. First a global search using genetic algorithms [6] and

next a local optimization using the steepest descent method [7]. Affine registration is also used as a pre-alignment step for non-rigid transformations described below.

- *Demons algorithm* is an intensity-based algorithm proposed by Thirion [8] 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 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. In the case of voxel-by-voxel intensity similarity, the instantaneous displacement vector for each voxel is

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

where f and g are the intensity image 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 [9]. In our case, parameter σ has been chosen by exhaustive search, between 0.6 and 2.0mm by steps of 0.2mm, minimizing the distance between the estimated STN and the real targets. Finally, σ of 1mm has been chosen.

- *B-splines algorithm*. It is a mutual information-based free-form deformation algorithm whose displacement field is modelled 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. in [10]. 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 has been implemented using the communication utilities for distributed memory architectures using the MPICH implementation of the Message Passing Interface (MPI) [11]. The good interpolation properties and the suitability for multiscale processing of the B-splines are well known [12] and its deformability can be controlled by changing the spacing between the control points of the grid which we have set at 12mm.

3 Results

3.1 Target Selection and Expert Variability

In order to evaluate the *repeatability* or *intra-expert* variability of the expert targeting, we have computed the centroid of each cloud of STN points targeted

Table 1. Expert variability statistics

(a) <i>Intra</i> -expert variability.			(b) <i>Intra</i> -expert variability. References.		
$m \pm s$	Expert 1	Expert 2	$m \pm s$	Left Ref.	Right Ref.
centroid	$1.06 \pm 0.61mm$	$0.80 \pm 0.52mm$	Exp. 1	$1.10 \pm 0.32mm$	$0.75 \pm 0.38mm$
			Exp. 2	$0.79 \pm 0.30mm$	$0.38 \pm 0.25mm$

by the expert and we have calculated the Euclidean distances from the centroid to each of these points, called *centroid variability*. In table 1(a) these statistics are shown for the two experts. These quantities allow us to get an idea of the surgeon variability and its accuracy when clicking over the pre-operative target. If we only consider the 2 STNs used as a reference we obtain a centroid variability that is shown in table 1(b) for the left and right sides respectively and for the different experts.

In order to construct our ground truth a paired *T-test* of the hypothesis that the target coordinates selected by each expert come from distributions with equal means has been performed over each coordinate (x, y, z) at a 1% significance level. The results show that the hypothesis can not be rejected. Therefore our reference can be considered as the mean of two cloud of points given by each expert. The *inter*-expert variability calculated as the Euclidean distances from each expert click to the ground truth gives a mean and unbiased standard deviation of $1.61 \pm 0.29mm$ and $1.40 \pm 0.38mm$ for the expert 1 and 2 respectively. These statistics have been obtained by generating one error per

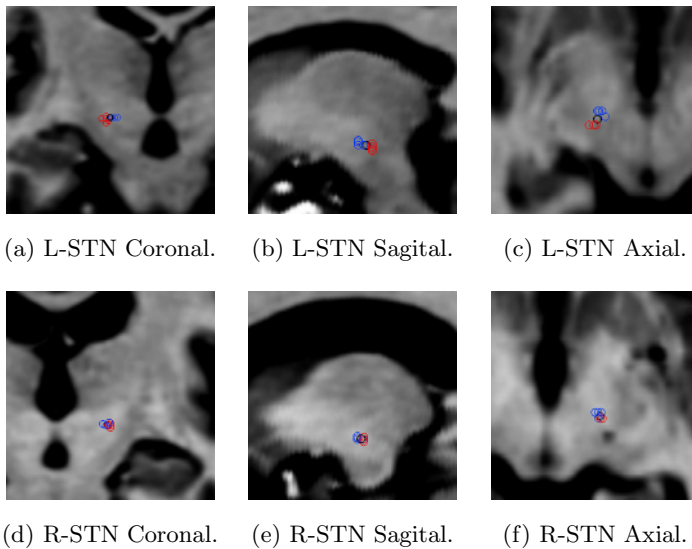


Fig. 1. Reference STN expert targeting

Table 2. Mean STN coordinates of the dataset and errors given by the methods

(a) STN coordinates referred to MCP.

Coordinates	<i>mean ± std</i>
AP	$-3.286 \pm 0.94mm$
LAT	$\pm 12.90 \pm 0.93mm$
VERT	$-3.23 \pm 0.78mm$
MCP-distance	$13.90 \pm 0.87mm$

(b) Estimation Errors.

Methods	<i>mean ± std</i>
Affine	$2.42 \pm 0.84mm$
Demons	$1.77 \pm 0.65mm$
B-Splines	$1.72 \pm 0.48mm$
AC-PC	$1.96 \pm 0.90mm$

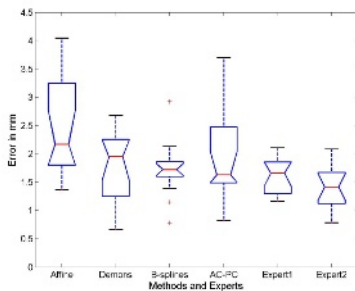
STN and per expert in the following manner. Each error is the Euclidean distance between the mean of the 5 points targeted by the expert and the ground truth.

In figure 1 the points targeted by the experts for the left (figures 1(a), 1(b), 1(c)) and right (figures 1(d), 1(e), 1(f)) STNs chosen as a reference can be seen. In blue the targeting of the expert 1, in red the targeting of the expert 2 and in black the mean point used as a reference. For visualization purposes, each point has been projected onto the three orthogonal planes passing through the centroid and showed using a circle (radius of 1mm).

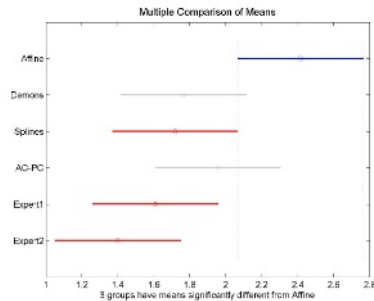
We can also decompose these distances in anteroposterior (AP), lateral (LAT) and vertical (VERT) coordinates which allows comparing directly the mean STN location for our database with usual STN coordinates from the atlases (e.g. [1]). In table 2(a) we show the mean and standard deviation of the coordinates referred to the MCP for the 16 STNs used in this study as well as their mean distance to the MCP.

3.2 Target Estimation and Evaluation of the Methods and Experts

The statistics, mean and unbiased standard deviation, of the errors committed when applying the 4 methods under study to locate the STN as described in 2.3 are shown in table 2(b). In order to compare the results statistical tests were performed. A one-way analysis of variance (ANOVA) of the hypothesis that the



(a) ANOVA statistical box plot.



(b) Multi-comparison test.

Fig. 2. Statistical tests of the errors committed using different methods and by the experts (using the *anova1* and *multcompare* functions of MATLAB[®])

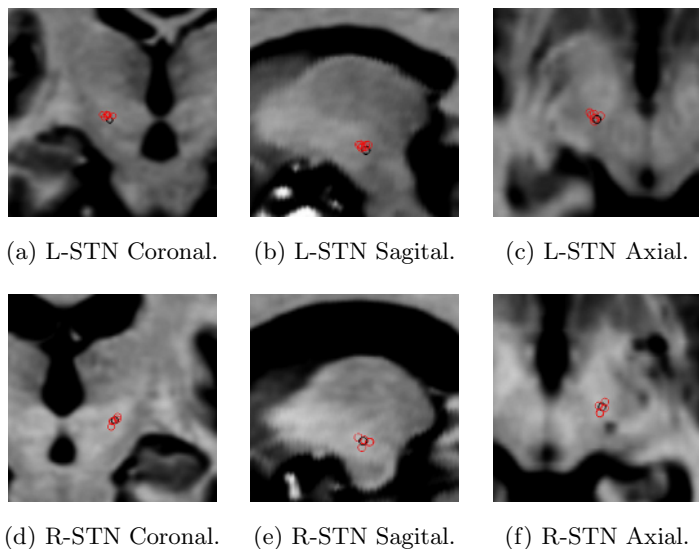


Fig. 3. STN estimation using B-splines

errors come from distributions with equal means has been performed over the errors produced by each method and by the experts at a 5% significance level. In figure 2(a) a statistical box plot produced by this test is shown as well as the result of a multi-comparison test of the means (see figure 2(b)). Two key conclusions can be drawn from these results. First that the errors committed with B-splines, demons and AC-PC based methods and the errors committed by the experts are not significantly different. Secondly that the results show that affine registration is significantly different from the B-splines method and from the experts. In figure 3 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 $1mm$ 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.

4 Discussion and Perspectives

The main conclusion one can extract from these results is that automatic STN location is possible and accurate. As we can see by simple inspection of the numerical results, the B-splines method shows the best performance with the smallest mean error and unbiased standard deviation but 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 while there are no significant differences between the errors committed using the B-splines, Demons or AC-PC referential-based techniques and, most importantly, that there are no significant differences between the errors committed with these three techniques and by the experts. Although the 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 inter-patient variability, which is very important at a single patient level. Although the choice of the STN of reference can influence the results, the methods have been tested using 6 different combinations of STN-pairs (left and right references) and the results were similar. The automatic estimation of the STN can also be used as a first and fast pre-operative target estimation that can be refined by the neurosurgeon criterion. In a near future, more registration methods, mainly local ones, and experts will be added to this study. Moreover, further work will include the construction of a post-mortem atlas of the STN which will provide a ground truth without the variability produced by expert targeting over a point.

References

1. Schaltenbrand, G., Wahren, W.: Atlas for Stereotaxy of the Human Brain. 2 edn. Thieme (1977)
2. Pollo, C. et al.: 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. Volume 80. (2003) 1–4
3. DHaes, P. F. et al.: Toward the creation of an electrophysiological atlas for the pre-operative planning and intra-operative guidance of deep brain stimulators (DBS) implantation. In: MICCAI. (2004) 729–736
4. Maes, F. et al.: Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging* **16** (1997) 187–198
5. Viola, P., Wells, W.M.: Alignment by maximization of mutual information. In: Proceedings of the Fifth ICCV, IEEE Computer Society (1995) 16
6. Wall, M.: GALib 2.4.6: A C++ Library of Genetic Algorithm Components. (Massachusetts Institute of Technology) <http://lancet.mit.edu/ga/>.
7. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical Recipes in C, The art of scientific computing. (Cambridge University Press)
8. Thirion, J.P.: Image matching as a diffusion process: an analogy with maxwell's demons. *MIA* **2** (1998) 243–260
9. Bach, M.: Atlas-based Segmentation and Classification of Magnetic Resonance Brain Images. PhD thesis, EPFL, CH-1015 (2003) Thesis Num. 2875.
10. Rueckert D. et al.: Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging* **18** (1999) 712–721
11. Gropp, W., Lusk, E., Skjellum, A.: Using MPI: Portable Parallel Programming With the Message-Passing Interface. MIT Press (1999)
12. Unser, M.: Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Mag.* **16** (1999) 22–38