

Evaluation and validation methods for intersubject non-rigid 3D image registration of the human brain

Ting Guo^{1,2}, Yves P. Starreveld³, and Terry M. Peters^{1,2,3}

¹Robarts Research Institute and University of Western Ontario,
London, Ontario, Canada N6A 5K8

²Biomedical Engineering Graduate Program, University of Western Ontario
London, Ontario, Canada N6A 5B9

³London Health Sciences Center, London, Ontario, Canada N6A 5A5
{tguo, tpeters} @ imaging.robarts.ca

ABSTRACT

This work presents methodologies for assessing the accuracy of non-rigid intersubject registration algorithms from both qualitative and quantitative perspectives. The first method was based on a set of 43 anatomical landmarks. MRI brain images of 12 subjects were non-rigidly registered to the standard MRI dataset. The “gold-standard” coordinates of the 43 landmarks in the target were estimated by averaging their coordinates after 6 tagging sessions. The Euclidean distance between each landmark of a subject after warping to the reference space and the homologous “gold-standard” landmark on the reference image was considered as the registration error. Another method based on visual inspection software displaying the spatial change of colour-coded spheres, before and after warping, was also developed to evaluate the performance of the non-rigid warping algorithms within the homogeneous regions in the deep-brain. Our methods were exemplified by assessing and comparing the accuracy of two intersubject non-rigid registration approaches, AtamaiWarp and ANIMAL algorithms. From the first method, the average registration error was 1.04mm +/- 0.65mm for AtamaiWarp, and 1.59mm +/- 1.47mm for ANIMAL. With maximum registration errors of 2.78mm and 3.90mm respectively, AtamaiWarp and ANIMAL located 58% and 35% landmarks respectively with registration errors less than 1mm. A paired t-test showed that the differences in registration error between AtamaiWarp and ANIMAL were significant ($P < 0.002$) demonstrating that AtamaiWarp, in addition to being over 60 times faster than ANIMAL, also provides more accurate results. From the second method, both algorithms treated the interior of homogeneous regions in an appropriate manner.

Keywords: non-rigid image registration, validation, landmark, visual inspection, brain mapping

1 INTRODUCTION

Planning functional neurosurgeries for the treatment of Parkinsonism, chronic pain, essential tremor, and some other movement disorders requires accurate localization of surgical targets for the placement of neurostimulators or lesions in specific regions in the midbrain or basal ganglia. Imaging techniques such as magnetic resonance imaging (MRI) and computerized tomography (CT) have been used to provide high-resolution images for preoperative planning and approximate localization. However, neither of them can provide sufficient contrast within the grey matter structures, particularly within the thalamus consisting of several functionally distinct subdivisions, which appears as a homogeneous region in most qualitative MR images. Due to the inadequacy of the information derived solely from these images, printed anatomical atlases [1,2] or digitized versions of them [3,4] have been adopted to assist localization of surgical targets that are otherwise indistinguishable in the images. Unfortunately, the striking variation of anatomical structures among different individual brains limits the applicability of printed atlases, since they are fixed and the accuracy of transferring the information on them to individual brains is largely dependent on the experience of the neurosurgeon. On the other hand the digitized atlases are extendable, flexible, and can be viewed on a computerized surgical planner on which atlases can be rigidly [5,6] or non-rigidly [7,9,12,28] aligned with individual preoperative brain images. Digitized atlases are therefore considerably more useful than their printed counterparts. Nevertheless the accuracy of existing anatomical atlases is not greatly reliable for stereotaxic neurosurgery because of limitations such as lack of morphometric information, poor generalization, insufficient statistical representation, non-uniform inter-slice

separation as well as brain tissue damage during the atlas construction process. To overcome the disadvantages described above, surgeons often employ electrophysiological techniques to refine the precision of the surgical target localization, to characterize tissue function, and to map somatotopy. Some research groups have been working on standardizing and normalizing the electrophysiological data from a series of patients to a standard reference space, such as CJH-27 dataset [24], to establish the relationship between brain functional organization and anatomic structures [26,9,13,19]. To facilitate the integration of electrophysiological data as well as digitized atlases into preoperative guidance procedures, rigid and non-rigid registration methods, accommodating intersubject anatomical variability, have been employed. In the last decade, Collins *et al.* [11,12], Schormann *et al.* [8], Thompson and Toga [21], Woods *et al.* [22,23], Starreveld [9,10] have contributed several intersubject registration algorithms to the literature.

In this paper, we presented methodologies to evaluate the performances of non-rigid intersubject registration algorithms and exemplified the approaches by assessing and comparing the accuracy of two non-rigid intersubject registration algorithms, the AtamaiWarp fast non-rigid registration algorithm [9,10] and the Automatic Non-rigid Image Matching and Anatomical Labeling (ANIMAL) algorithm [11,12]. Previously, a number of methods have been published for describing the problem of non-rigid intersubject registration validation, including direct visual assessment, the overlay of manually and automatically segmented structures [12], measurement of distances between corresponding landmarks [14,20], overlap of structures [15], overlap of segmented structures [16], image subtraction [17], and maps of three-dimensional variability [18]. Each method has some significant shortcomings for application to the validation of non-rigid deformation registration of the deep-brain. Percentage overlap of structures and distances between segmented surfaces both require segmentation of the structures to be assessed, which itself requires a significant amount of processing. Furthermore, in the area of the deep-brain, the structures are small and poorly defined. Image subtraction is sensitive to intensity variations inherent to the acquisition and may falsely indicate misregistration. Considering both quantitative and visual inspecting perspectives, we developed our validation methods based on anatomical landmarks and an in-house 3D visualization environment. Although landmark-based studies are difficult to perform, because their selection relies on subjective judgments, the unique properties of the image and brain structures they represent are important in validation of non-rigid registration procedures. Hence we wanted to use anatomical landmarks to evaluate qualities of global transformations and local alignments of both the AtamaiWarp and ANIMAL non-rigid intersubject registration approaches. Another critical point that concerned us was how to evaluate the performance of an intersubject non-rigid registration algorithm within homologous regions bounded by edges. The degree of correspondence of anatomical landmarks can tell us how well homologous landmarks and edges align, but it does not provide information about how well the algorithms behave within homogeneous regions. Clinically, most of the surgical targets for the treatment of Parkinson's disease are located inside homogeneous regions such as the globus pallidus, subthalamic nucleus, and motor thalamus. It is therefore especially critical to understand the behaviour of the non-rigid registration algorithms within these regions. To address this issue we developed an intuitive visualization approach, which allows us to place blocks of colour-coded spheres within such homogeneous regions. The resulting distortion of the cube geometry after being subject to the non-rigid transformation matrix generated by the registration allows us to visualize the effect of the algorithm within homogeneous regions. By comparing the overlap between the colour-coded spheres and the 3D segmentation of a digitized Schaltenbrand and Wahren atlas [1] before and after deformation, we were also able to examine the performance of each algorithm quantitatively.

2 METHODS

2.1 Landmark-based validation method

2.1.1 Subjects

In this evaluation, the MRI scans of 12 subjects (11 male, 1 female) with an average age of 57 were randomly selected from a database containing data of 88 patients who had undergone surgeries for symptomatic treatment of Parkinson's disease, chronic pain and essential tremor at London Health Sciences Centre (LHSC), London, Ontario, Canada. Eight of these 12 subjects were diagnosed with Parkinson's disease and four with essential tremor. Three of them had a left pallidotomy, four a right thalamotomy and five a left thalamotomy.

2.1.2 MRI data

MRI data of the patients were acquired with a 1.5T GE Signa scanner using a 3D SPGR sequence (TR/TE 8.9/1.9, flip angle 20°, NEX 2). The in-slice resolution was 256×256 and the voxel size was $1.17\text{mm} \times 1.17\text{mm} \times 1\text{mm}$.

2.1.3 Standard reference brain template

CJH-27 dataset [24] was adopted as the reference MRI brain template (the warping target) because of the fine anatomical detail it provides and the high SNR it has. CJH-27 consists of 27 registered T_1 – weighted MRI scans ($20 \times 1\text{mm}^3$: TR/TE 18/10ms, flip angle 30°, NEX 1; $7 \times 0.78\text{mm}^3$: TR/TE 20/12ms, flip angle 40°, NEX 1) of the same healthy individual averaged into a single volume.

2.1.4 Visualization platform

To facilitate anatomical landmark selection, a computerized visualization and development environment, the Atamai Surgical Planner (ASP) platform [9], was employed in our validation protocols. Developed by Finnis et al. in the Imaging Research Laboratories at the Robarts Research Institute, ASP can interactively display stereoscopic 3D image data comprising of three intersecting orthogonal image planes (axial, coronal and sagittal), which are able to be moved through the whole image volume respectively and resliced at oblique angles.

2.1.5 Non-rigid intersubject registration algorithms

The ANIMAL algorithm [11,12] was developed by Collins et al. at the Montreal Neurological Institute for intersubject MRI registration. It uses a multi-resolution strategy to minimize a cost function that values the normalized cross correlation of the images and avoids large deformations. The algorithm eventually produces a three dimensional displacement grid specifying a transformation from a source volume to a target volume. Although the intersubject registration quality of ANIMAL algorithm was excellent by visual assessment, it typically took 8 hours on a typical hardware platform (SGI Origin R10K @ 225Mhz) to warp an entire subject brain to the reference image volume using a final grid spacing of 2mm.

Inspired by Collins' ANIMAL algorithm, Starreveld developed the AtamaiWarp algorithm [10], also a multi-resolution approach, but built using standard VTK classes, and optimized to run a specific hardware, with an execution time of 8 to 12 minutes on a dual PIII 933 MHz machine. Rather than warping the entire brain, AtamaiWarp can target the registration of structures in the deep-brain region enclosed within a user-defined mask. Primarily this registration methodology consists of two separate steps, the generation of a global affine transformation and the computation of a deformation grid building upon the first step. The affine transformation maximizes the normalized cross-correlation between the source and target volumes, and the deformation grid maximizes a similarity metric on successively smaller cubic sub-volumes of the images. The output of this algorithm is a 3D displacement grid with a space of 2mm that specifies the transformation from an input volume to the reference image.

Both algorithms are completely unsupervised. This advantage eliminates the subjective error introduced by manually selecting landmarks for establishing the transformations.

2.1.6 Anatomical landmark identification

This validation methodology is based on a set of 43 anatomically distinct landmarks, including anterior commissure and posterior commissure, allowing evaluation of the non-rigid transformation accuracy. The 43 landmarks were repeatedly identified in both the reference CJH-27 template and also within each subject image using ASP to assist in visualization. In our work, each selected anatomical landmark was represented by an individual sphere with a user-specified colour in 3D image space. We list the defined 43 landmarks in **Table 1**, as well as their “gold standard” coordinates. The following criteria were adopted for the landmark selection: (1) Simple, quick and reproducible identifiability; (2) Stereoscopic identification on three intersecting orthogonal planes (axial, coronal and sagittal); (3) Characterization of the topographic uniqueness of deep-brain regions.

Estimation of the “gold-standard”: An individual rater performed the localization of the 43 landmarks in the standard CJH-27 dataset 6 times during five-week period in which tagging sessions were spaced 1 week apart. We applied an analysis of variance (ANOVA) on the coordinates of these 6 sessions to determine the amount of intrarater variability and to determine whether the variability of between groups is significant. The “gold-standard” coordinates of these 43

landmarks were estimated by averaging their coordinates of these 6 tagging sessions. To examine the reliability of the “gold-standard”, we calculate the root mean squared (RMS) error of the landmark localization and standard deviation of localizing landmarks in x, y, and z directions in 6 sessions.

Landmarks		CJH-27 dataset			Registration Error (mm)	
		x	y	z	Atamai	ANIMAL
01	Frontal horn of lateral ventricle anterior L	17.542	30.646	0.686	0.44	1.80
02	Frontal horn of lateral ventricle anterior R	19.014	29.674	0.994	0.36	0.82
03	Head of caudate nucleus anterior L	-18.576	24.684	0.362	0.41	1.10
04	Head of caudate nucleus anterior R	20.382	25.274	0.232	0.25	0.83
05	Putamen anterior L	-21.720	19.838	-4.212	0.31	0.68
06	Putamen anterior R	23.778	21.176	-2.130	0.93	1.76
07	Anterior commissure	0.466	2.648	-4.752	0.46	1.08
08	Posterior commissure	0.206	-23.822	-3.196	0.18	0.54
09	Occipital horn of lateral ventricle L	-19.342	-41.076	11.258	0.93	3.60
10	Occipital horn of lateral ventricle R	19.554	-38.104	11.634	1.06	3.65
11	Lateral ventricle posterior inferior L	-4.836	6.486	-4.154	0.84	1.20
12	Lateral ventricle posterior inferior R	5.326	8.092	-4.018	0.05	0.41
13	Putamen posterior L	-29.860	-21.414	2.508	0.59	2.79
14	Putamen posterior R	29.716	-20.370	3.564	0.69	1.12
15	Putamen medial L	-13.866	9.132	-4.962	0.42	0.76
16	Putamen medial R	15.180	10.626	-4.912	0.77	0.94
17	Frontal horn of lateral ventricle superior L	-18.124	-25.420	31.710	2.78	1.61
18	Frontal horn of lateral ventricle superior R	17.570	-23.818	30.878	2.43	0.63
19	Caudate nucleus lateral superior L	-20.966	-1.058	19.024	1.09	3.90
20	Caudate nucleus lateral superior R	21.132	-0.582	19.170	1.20	2.01
21	Putamen superior L	-26.710	-1.724	15.234	1.17	1.61
22	Putamen superior R	27.548	-0.392	13.722	1.19	1.26
23	Caudate nucleus superior L	-18.634	-5.602	25.488	2.63	0.70
24	Caudate nucleus superior R	18.400	-7.682	24.086	1.75	2.09
25	Globus pallidus, pars lateralis superior L	-19.674	-2.904	5.942	0.93	1.90
26	Globus pallidus, pars lateralis superior R	19.998	-2.206	5.724	0.73	0.95
27	Temporal horn of lateral ventricle lateral L	-39.067	-19.394	-16.026	1.56	1.28
28	Temporal horn of lateral ventricle lateral R	41.498	-17.800	-17.962	1.63	1.40
29	Globus pallidus, pars lateralis medial L	-9.100	3.562	-5.364	0.43	1.13
30	Globus pallidus, pars lateralis medial R	10.148	3.450	-4.512	0.60	1.54
31	Corpus callosum-anterior tip of genu	0.504	35.612	2.060	0.75	0.96
32	Interthalamic adhesion medial	0	-12.592	5.098	0.82	2.02
33	Mamillary body L	-1.589	-6.345	-16.703	1.48	2.74
34	Mamillary body R	3.118	-6.156	-16.810	1.52	0.72
35	Clastrum putamen adhesion L	-32.894	0.506	-2.620	0.51	2.89
36	Clastrum putamen adhesion R	34.492	0.336	-2.616	0.22	1.99
37	Hippocampus inferior L	-22.676	-8.770	-28.592	2.43	3.81
38	Hippocampus inferior R	26.730	-7.732	-28.050	2.35	3.43
39	Putamen inferior L	-27.528	-1.006	-12.928	1.24	1.57
40	Putamen inferior R	29.860	-1.104	-11.790	1.14	0.44
41	Corpus callosum-posterior tip of genu	1.078	23.346	4.348	0.47	0.31
42	Corpus callosum fornix junction inferior anterior	0.308	-14.600	22.862	2.03	0.63
43	Corpus callosum fornix junction superior anterior	0.930	11.886	-0.688	0.99	1.83

Table 1. Definitions of the 43 landmarks and their “gold-standard” coordinates.

Preprocessing: Each subject image was non-rigidly registered to the reference CJH-27 brain template using a non-rigid intersubject registration algorithm to establish the 3D transformations and deformation grids. In this study we acquired the transformations and the displacement vector files with ANIMAL and AtamaiWarp algorithms respectively. Each resulting transformation matrix and deformation grid was then used to map the landmarks in each corresponding subject image to the reference brain volume.

Transforming landmarks: To be compared with the 43 “gold-standard” landmarks in CJH-27 dataset, each set of landmarks in a subject space was non-rigidly registered to the reference brain space according to a specific displacement vector file generated by warping each subject image to the CJH-27 template, using each algorithm in turn.

2.1.7 Registration accuracy analysis

The registration error in our study was defined as the Euclidean distance between each landmark of a subject after non-rigid registration and the “gold-standard” homologous landmark in the CJH-27 template. After each non-rigid registration operation, the resulting transformation was applied to the coordinates of each of the 43 landmarks in each subject in turn. The differences of the resulting coordinates among 12 subjects were analyzed with one-way ANOVA to reveal the differences among them. The mean location of each landmark across 12 subjects after warping with each algorithm was calculated and compared with that of the corresponding homologous “gold-standard” landmark to generate the registration error. The registration accuracy of each algorithm was assessed by evaluating the registration errors of 43 landmarks. Differences between the registration accuracies of AtamaiWarp and ANIMAL were analyzed using a paired t-test.

2.2 Visual inspection validation method

2.2.1 Visual inspection display

The performance of each algorithm in homogeneous regions such as thalamus, putamen and globus pallidus, was subjectively analyzed using a visualization and evaluation environment programmed using VTK and Python. The objective here was to provide a means of easily visualizing the non-rigid transform that results within a homogeneous region. The transform itself is driven by features surrounding such regions, whereas the effect of the transformation within such regions is not evident. However since target locations may fall inside such regions, it is important to understand how a non-rigid registration algorithm handles such target points. A well-behaved algorithm would be expected to treat the interior of such region as if it were the 3D equivalent of a rubber sheet. On the other hand, if points in the center of such region ended up close to the peripheral structures of the homogeneous region in the target image after registration, we would consider that the algorithm failed in such cases. In order to visualize the behaviour of AtamaiWarp and ANIMAL algorithms in such regions, we construct a sphere-filled cube, which could be interactively placed at arbitrary position in the source image. If the warping grid produced by either non-rigid registration algorithm is applied to this cube, the new distribution of the colour-coded spheres will reflect the behaviour of the algorithm at the region occupied by the cube. These spheres are colour-coded to ensure that their ordering after transformation can be compared with that established a-priori.

2.2.2 Overlap with segmented nuclei

The digitized Schaltenbrand atlas described by St. Jean et al [27] demonstrates the parcellation of the various sub-nuclei of the thalamus. We employed the dataset to assess the registration accuracy of the non-rigid warping algorithms inside of homogeneous regions within the deep-brain. For each algorithm being evaluated, the digitized atlas and several segmented nuclei within the CJH-27 template were deformed to each subject brain space, according to the inverse of each established transformation acquired by warping each subject brain image to the reference CJH-27 dataset. The sphere-filled cubes placed in the regions across Vce and Vci and across Vim and Ce of the thalamus in CJH-27 were similarly treated. By comparing the number of spheres contained in each nucleus in CJH-27 dataset and that contained in the homogeneous nucleus in each patient brain after registration with each algorithm, we were able to evaluate the accuracy of both the ANIMAL and AtamaiWarp algorithms within homogeneous regions in the deep-brain. For these tests we employed the same MRI datasets as those used in the landmark-based validation method.

3 RESULTS

3.1 Landmark-based validation results

3.1.1 “Gold-standard” coordinates of 43 landmarks

43 landmarks in the standard brain template were repeatedly tagged 6 times with one-week interval between each tagging session. To examine the intrarater differences in the “gold standard” landmark localization, we performed three one-way ANOVAs on the landmark coordinates across the 6 sessions. There were six groups (six tagging sessions) for each ANOVA. In each group, the number of samples equals to the number of landmarks and each sample being the coordinate of either x, y, or z. The results showed that the intersession variance was not significant in x (df = 5, F = 0.0004, P = 1), y (df = 5, F = 0.0024, P > 0.9999), and z (df = 5, F = 0.0034, P > 0.9999) directions. Significant correlation was indicated in the Pearson-Product Moment correlation study of each of the three dataset ($r > 0.99$). The RMS error of localizing these points is 0.70mm and the average standard deviation was 0.37mm, 0.78mm and 0.64mm in x, y and z directions respectively. These results demonstrated that our “gold-standard” coordinates are sufficient reliable to act as a reliable reference.

3.1.2 Registration accuracy

One-way ANOVAs were performed to detect whether the resulting coordinates of each of the 12 subjects after transformation with an algorithm differ significantly from one another. The results showed that the transformed coordinates of the 12 subjects are not significantly different in x, y, and z directions for either ANIMAL (df = 11, F < 0.0585, P > 0.99) or AtamaiWarp algorithm (df = 11, F < 0.0040, P = 1). Therefore, after averaging the coordinates of each homologous landmark across 12 subjects after registration with each algorithm and comparing them with those of the corresponding “gold standard” landmark, we obtained that the average registration error was 1.04mm +/- 0.65mm for AtamaiWarp, and 1.59mm +/- 1.47mm for ANIMAL. The maximum, minimum, average registration errors and their standard deviations over the database for the tested algorithms are given in **Table 2**.

		AtamaiWarp	ANIMAL
Registration Error (mm)	Average	1.04	1.59
	Maximum	2.78	3.89
	Minimum	0.05	0.31
Sd of Registration Error (mm)	Average	0.65	1.47
	Maximum	1.88	5.06
	Minimum	0.13	0.32

Table 2. Average, maximum, and minimum registration errors and standard deviations.

3.1.3 Comparison of AtamaiWarp & ANIMAL

AtamaiWarp and ANIMAL, with maximum registration errors of 2.78mm and 3.90mm respectively, located 58% and 35% of the landmarks respectively with registration errors less than 1mm. The paired t-test performed to assess the difference with respect to registration error between AtamaiWarp algorithm and ANIMAL algorithm demonstrated that these two non-rigid intersubject registration algorithms are significantly different ($t = 3.34$, df = 42, P < 0.002), and AtamaiWarp, in addition to being over 60 times faster than ANIMAL, also provides more accurate results. In addition, 43 paired t-tests comparing the registration error of each landmark across 12 subjects between the two algorithms were also conducted to verify our conclusion. In 30 out of 43 landmarks there were significant differences between AtamaiWarp and ANIMAL algorithms ($t > 2.26$, df = 11, P < 0.045).

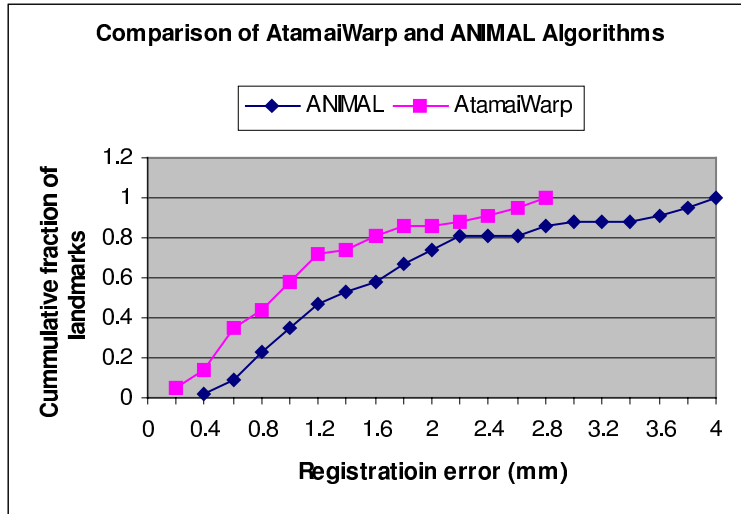


Figure 1. Comparison of AtamaiWarp and ANIMAL algorithms in registration error, clearly indicating the superiority of AtamaiWarp over ANIMAL. AtamaiWarp registered much more amount of landmarks with less registration errors.

3.2 Visual inspection results

In **Figure 2-a**, a colour-coded sphere-filled cube is placed in thalamus region of a subject brain image. **Figure 2-b** displays the new distributions of those spheres after warping using AtamaiWarp algorithm in the CJH-27 brain template, and those after warping using ANIMAL algorithm in CJH-27 are shown in **Figure 2-c**. Although the results of both algorithms are slightly different from each other, they both align the embedded spheres in the expected manner. Consistent with our expectation, both algorithms treat the interior of homogeneous regions as 3D elastic-equivalent material. Note the well-behaved “stretching” of the cube, reflecting the influence of the structures at the periphery of the thalamus and its interior wall. The multi-colour spheres inside the thalamus in the subject image are moved to approximately homologous locations of the homogeneous region in the reference image after registration by either algorithm. The colour-order in each of the three directions remains unchanged, indicating that isomorphic mappings are preserved.

When comparing the overlap between the colour-coded spheres and the 3D segmentation of digitized Schaltenbrand and Wahren atlas before and after deformation (**Figure 3**), we found that although slight differences again exist between the results of AtamaiWarp and those of ANIMAL, they are not significant. Since the preservation of isomorphic mappings of the colour-coded spheres were demonstrated, and the average number of spheres within each nucleus in each patient brain registered by either algorithm is very close to that in CJH-27 template before registration as shown in **Table 3**, the results of this method indicate excellent registration accuracy of both algorithms within homogeneous regions in deep-brain.

	CJH-27	AtamaiWarp	ANIMAL
Vce	50	50.36	50.18
Vci	54	50.55	50.72
Ce	34	30.09	32.64
Vim	28	26.36	27.91

Table 3. The average number of spheres contained in each nucleus before and after warping.

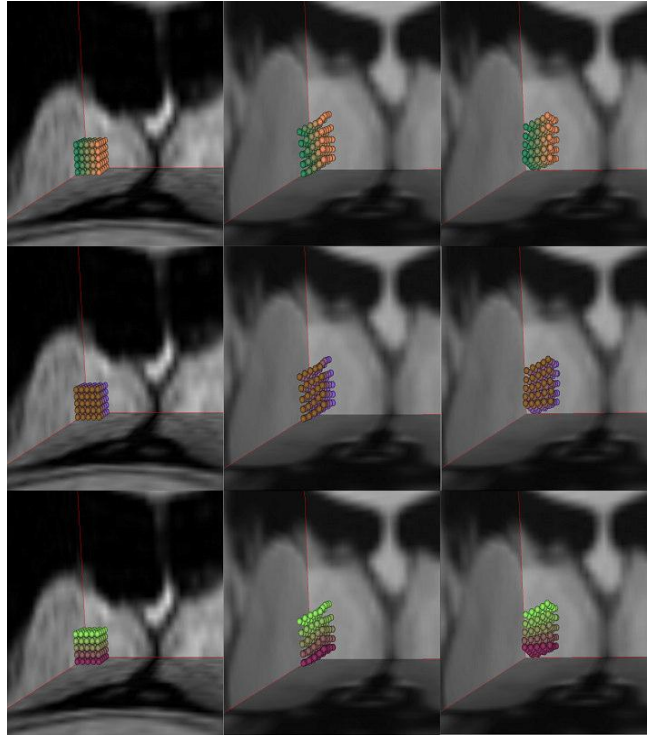


Figure 2. a. Left: multi-colour spheres in a subject brain image; **b.** Middle: homologous spheres after non-rigid intersubject registration in reference CJH-27 template using AtamaiWarp algorithm; **c.** Right: using ANIMAL algorithm. Colour changes along three directions. This test demonstrated that while the performance of each algorithm differs slightly, both are “well-behaved”.

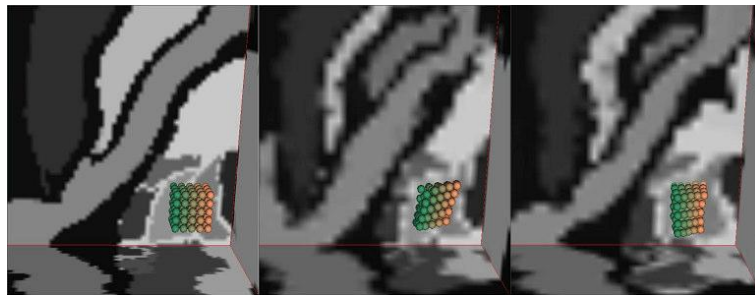


Figure 3. Left: colour-coded spheres enclosed within segmented Vce and Vci nuclei of thalamus in CJH-27 brain template; **Middle:** spheres and segmented nuclei were non-rigid registered to a subject brain image using AtamaiWarp algorithm; **Right:** using ANIMAL algorithm.

4 DISCUSSION

Non-rigid intersubject registration has been widely used to construct deformable digital atlases [25] and 3D electrophysiological databases [9] for neurosurgery planning and guidance. The effectiveness of the atlases and databases is highly depending on the accuracy of these algorithms. As there is no “ground truth” for measuring their accuracy in a clinical situation, we have chosen a landmark-based approach to assess two non-rigid intersubject registration algorithms and have compared this with visual assessment approach. Our landmark-based approach was

performed within a visualization environment capable of interactively displaying 3D image data in arbitrary intersecting orthogonal image planes. Our approach differed from other landmark-based methods [14,20], in that our method localized each landmark on three orthogonal planes to ensure its uniqueness. The anatomical landmarks involved in our study are well and clearly defined and can be located quickly. Using this approach, the gold standard demonstrated high reliability on 6 repeated measurements, which was proved by both RMS error and Pearson correlation studies. The RMS error is also referred to as intrarater reliability; meanwhile correlation provided the intrarater consistency. Results from different statistical analyses employed in this study even have strong agreement. With the assistance of our novel visual inspection system, we were also able to assess the performance of both algorithms within some deep-brain structures. Our results demonstrated that although both AtamaiWarp and ANIMAL algorithms performed well in non-rigid intersubject registration, the AtamaiWarp algorithm is more accurate than ANIMAL. This landmark-based method and the visual inspection environment can be applied to validate other non-rigid registration algorithms, and offer opportunity to examine non-rigid registration algorithms at subvoxel level. We note that some excellent evaluation methods [27,29] for intersubject brain registration are available, however they focus more on cortex regions rather than deep-brain. Our approaches therefore complement existing techniques and permit the evaluation of intersubject non-rigid 3D image registration in the deep brain.

There are four options for performing non-rigid registration available in Insight Segmentation and Registration Toolkit (ITK), Demons, FEM-Based, BSplines, and LevelSet techniques. A non-rigid registration of brain images of size 256 x 256 x 200 can typically take one hour to execute on a modern desktop machine (e.g. Pentium 4 at 2 GHz) [30]. Although the computation time can be reduced dramatically by providing reference landmarks or performing an affine registration before the non-rigid one, we primarily concerned here with the speed of the unsupervised non-rigid registration.

In this study, the CJH-27 dataset was adopted as the standard anatomic template and each subject brain image was non-rigidly registered to it with AtamaiWarp and ANIMAL algorithms. As CJH-27 is based on registered and averaged MRI scans of the same healthy individual aged 27 [24], which might not be an ideal reference in registration of images of aged patients with Parkinson's disease, it is important to consider that the future studies should employ a more suitable template derived from multiple images of age-matched patients.

REFERENCES

1. G. Schaltenbrand and W. Wahren, *Atlas for Stereotaxy of the Human Brain*. Stuttgart, Germany: Thieme, 1977.
2. J. van Buren and R. Borke, *Variations and Connections of the Human Thalamus*. Berlin, Germany: Springer, 1972.
3. W. L. Nowinski, T. T. Yeo, and G. L. Yang, "Atlas-based system for functional neurosurgery," in *Proc. SPIE Medical Imaging*. Bellingham, WA: SPIE, 1997, vol. 3031, Image Display, pp. 92-103.
4. G. Bertrand, A. Olivier, and C. J. Thompson, "The computerized brain atlas: Its use in stereotaxic surgery," *Trans. Amer. Neurological Assoc.*, vol. 98, pp. 233-237, 1973.
5. B. A. Kall, P. J. Kelly, S. J. Goerss, and G. Frieder, "Methodology and clinical experience with computed tomography and a computer-resident stereotactic atlas," *Neurosurgery*, vol. 17, no. 3, pp. 400-407, 1985.
6. R. M. Lehman, J. Zheng, J. L. Hamilton, and E. Micheli-Tzanakou, "Comparison of 3-D stereoscopic MR imaging with pre and post lesion recording in pallidotomy," *Acta Neurochir. (Wein)*, vol. 142, pp. 319-328, 2000.
7. P. St-Jean, A. F. Sadikot, D. L. Collins, D. Clonda, R. Kasrai, A. C. Evans, and T. M. Peters, "Automated atlas integration and interactive 3-dimensional visualization tools for planning and guidance in functional neurosurgery," *IEEE Trans. Med. Imag.*, vol. 17, pp. 672-680, Oct. 1998.
8. T. Schormann, S. Henn, and K. Zilles, "A new approach to fast elastic alignment with application to human brains," *Lecture Notes Comput. Sci.*, vol. 1131, pp. 437-442, 1996.
9. K. W. Finnis, Y. P. Starreveld, A. G. Parent, A. F. Sadikot, and T. M. Peters, "Three-Dimensional Database of Subcortical Electrophysiology for Image-Guided Stereotactic Functional Neurosurgery," *IEEE Trans. Med. Imag.*, vol. 22, pp. 93-104, Jan. 2003.
10. Y. P. Starreveld, "Fast non-rigid registration applied to stereotactic functional neurosurgery," Ph.D. dissertation, Univ. Western Ontario, London, ON, Canada, June 2002.
11. D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," *J. Comput Assist Tomogr*, vol. 18, pp. 292-305, 1994.
12. D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans, "Automatic 3-D modelbased neuroanatomical segmentation," *Human Brain Mapping*, vol. 3, pp. 190-208, 1995.
13. G. Bertrand, J. Blundell, and R. Musella, *Electrical Stimulation of the Internal Capsule and Neighboring Structures During Stereotactic Procedures*. Philadelphia, PA: Harvey Cushing Soc., Apr. 20, 1963.

14. I. D. Grachev, D. Berdichevsky, S. L. Rauch, S. Heckers, D. N. Kennedy, V. S. Caviness, and N. M. Alpert, "A method for assessing the accuracy of intersubject registration of the human brain using anatomic landmarks," *Neuroimage*, vol. 9, pp. 250-268, 1999.
15. G. E. Christensen and H. J. Johnson, "Consistent image registration," *IEEE Trans Med. Imag.*, vol. 20(7), pp. 568-582, 2001.
16. J. A. Fiez, H. Damasio, and T. J. Grabowski, "Lesion segmentation and manual warping to a reference brain: intra- and interobserver reliability", *Human Brain Mapping*, vol. 9, pp. 192-211, 2000.
17. J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackal, A. D. Castellano-Smith, T. Hartkens, G. P. Penny, W. A. Hall, L. Haiying, C. L. Truwit, F. A. Gerritson, D. G. Hill, and D. J. Hawkes, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," *Proc MICCAI 2001*, W. Neissen and M. Viergever (Eds.), 572-581, 2001.
18. A. W. Toga and P. M. Thompson, "The role of Image registration: In brain mapping," *Image and Vision Computing*, vol. 19, pp. 3-24, 2001.
19. E. G. Duerden, K. W. Finnis, T. M. Peters, and A. F. Sadikot, "A method for analysis of electrophysiological responses obtained from the motor fibers of the human internal capsule," *Proc MICCAI 2003*, R. E. Ellis and T. M. Peters (Eds.) vol. 2, pp. 50-57, 2003.
20. S. Arndt, R. Rajarethinam, T. Cizadlo, D. O'Leary, J. Downhill, and N. C. Andreasen, "Landmark-based registration and measurement of magnetic resonance images: A reliability study," *Psychiatry Res. Neuroimaging*, vol. 67, pp. 145-154, 1996.
21. P. M. Thompson, and A. W. Toga, "A surface-based technique for warping three-dimensional images of the brain," *IEEE Trans. Med. Imaging*, vol. 15(4), pp. 402-417, 1996.
22. R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta, "Automated image registration: I. General methods and intrasubject, intramodality validation," *J. Comput. Assist. Tomogr.*, vol. 22(1), pp. 139-152, 1998.
23. R. P. Woods, S. T. Grafton, J. D. Watson, N. L. Sicotte, and J. C. Mazziotta, "Automated image registration: II. Intersubject validation of rigid and non-rigid models," *J. Comput. Assist. Tomogr.*, vol. 22(1), pp. 153-165, 1998.
24. C. J. Holmes, R. Hoge, D. L. Collins, R. Woods, A. W. Toga, and A. C. Evans, "Enhancement of MR images using registration for signal averaging," *J. Comput. Assist. Tomogr.*, vol. 22, no. 2, pp. 324-333, 1998.
25. K. A. Ganser, H. Dickhaus, R. Metzner, and C. R. Wirtz, "A deformable digital brain atlas system according to Talairach and Tournoux," *Medical Image Analysis*, vol. 8(1), pp. 3-22, 2004.
26. M. Yoshida, "Three-dimensional electrophysiological atlas created by computer mapping of clinical responses elicited on stimulation of human subcortical structures," *Stereotact. Funct. Neurosurg.*, vol. 60(1-3), pp. 127-134
27. P. Hellier, C. Barillot, I. Corouge, 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, pp. 1120-1130, Sep. 2003.
28. P. St-Jean, A. F. Sadikot, L. Collins, D. Clonda, R. Kasrai, A. C. Evans, and T. M. Peters, "Automated Atlas Integration and Interactive Three-Dimensional Visualization Tools for Planning and Guidance in Functional Neurosurgery," *IEEE Trans. Med. Imag.*, vol. 17, pp. 672-680, Oct. 1998.
29. S. Robbins, A. C. Evans, D. L. Collins, and S. Whitesides, "Tuning and Comparing Spatial Normalization Methods," *Medical Image Analysis*, vol. 8(3), pp. 311-323, 2004.
30. L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *ITK Software Guide*. New York, USA: Kitware, Inc., 2003.