

Retrospective Evaluation of Inter-subject Brain Registration

P. Hellier¹, C. Barillot¹, I. Corouge¹, B. Gibaud², G. Le Goualher^{2,3}, L. Collins³,
A. Evans³, G. Malandain⁴, and N. Ayache⁴

¹ Projet Vista, IRISA/INRIA-CNRS Rennes, France

² Laboratoire SIM, Hôpital de Pontchaillou, Rennes, France

³ Montreal Neurological Institute, Mc Gill University, Canada

⁴ Projet Epidaure, INRIA Sophia-Antipolis, France

Abstract. Although numerous methods to register brains of different individuals have been proposed, few work has been done to evaluate the performances of different registration methods on the same database of subjects. In this paper, we propose an evaluation framework, based on global and local measures of the quality of the registration. Experiments have been conducted for 5 methods, through a database of 18 subjects. We focused more extensively on the registration of cortical landmarks that have a particular relevance in the context of anatomical-functional normalization. For global measures, results show that the quality of the registration is directly related to the transformation's degrees of freedom. However, local measures based on the matching of cortical sulci, did not make it possible to show significant differences between affine and non linear methods.

Key words: Evaluation, non-rigid registration, atlas matching, neuroanatomy, MRI, cortical sulci.

1 Introduction

The comparison of brains of different individuals is an ancient objective in medicine. It has been pursued for a long time and was traditionally treated by paper-based atlases, with generally rather simple transformations. However, during the last few years, the development of electronic brain atlases [4,10,14] has emerged by overcoming some limitations of traditional paper-based atlases. To build such an atlas, it is necessary to compare brains of different individuals, so that each new subject contributes to the evolution and the relevance of the atlas. The comparison of brains requires the development of a registration method, most often with a non-rigid transformation.

An increasing number of authors study this registration problem. As it would be a gargantuan task to quote them all, we refer the reader to [9] for an overall survey on that subject. These methods are generally divided into two groups: intensity-based methods, that rely generally on the matching of voxels having comparable luminance (for mono-modal registration), and feature-based methods that rely on the extraction and matching of sparse landmarks. Feature-based methods dramatically depend on the extraction of features, and are generally valid near these features. In contrast, "photometric" methods use the entire available information, and make it possible to estimate

transformations with high degrees of freedom. This simple comparison may explain the popularity of intensity-based methods, which has been proved in the particular context of rigid multimodal fusion [16].

Nevertheless, the superiority of "iconic" methods has not been proved in the context of mono-modality inter-individual fusion. As a matter of fact, these methods usually rely on the minimization of an appropriate cost function, that exploits a relationship between voxels' luminance. The different methods mainly differ by the regularization scheme, and by the optimization strategy, which have a crucial consequence on the registration process. Published methods manage to optimize the matching criterion, but we do not really know if the formulation of the problem, combined with the way it is solved, lead to anatomically consistent transformations. Is it relevant to deform one subject toward another? What can we expect from the different registration methods? These questions are the starting point and the motivation of our work.

The evaluation project was conducted for 5 registration methods on a database of 18 subjects. The Vista project (INRIA-CNRS, Rennes) gathered the registration results, i.e. the deformations fields that were used to deform specific anatomical landmarks. The goal of this project is to evaluate how anatomical features are matched by the registration methods.

The paper is organized as follows : section 2 presents briefly the methods that have been evaluated, section 3 presents the data and the evaluation criteria which were used in the evaluation project. Section 4 details the results on a database of 18 subjects with global and local evaluation of the registration methods. Conclusions are drawn in section 5.

2 Participants

This evaluation project is somehow inspired by the Vanderbilt evaluation project [16], since all participants downloaded the data and performed the registration processes in their own laboratory. The results, i.e. the deformation fields, were then sent to our group (Vista Project, IRISA) to be evaluated on the basis of criteria that were not available to the participants of the evaluation project.

So far, 5 methods have been evaluated. We do not describe extensively the different methods, referring the reader to adequate references. We have adopted the following denomination for the methods:

- Method A. The denomination refers to the ANIMAL algorithm developed by L. Collins *et al.* at the MNI [2]. It must be noted that the finest resolution of the method A is 4 mm, for which the deformation field is piecewise constant.
- Method D. The denomination refers to the Demon's algorithm developed by J.P. Thirion in the Epidaure Group at INRIA Sophia-Antipolis [13].
- Method M. This registration is a simple rigid transformation, obtained by maximization of mutual information [8,15]. Although inadequate in that context, this method was implemented as a comparison basis for non-rigid methods.
- Method P. This method is the proportional squaring of Talairach. The method is based on the identification of the points AC-PC, which define a piecewise affine transformation on 12 cubes [12].

- Method R. The method R was developed at INRIA Rennes by P. Hellier *et al.* [7]. It may of course be questionable that the authors of an evaluation project submit their own registration method to the evaluation. Despite this, we hope that the reader would believe that we acted faithfully.

3 Data and Evaluation Criteria

For the evaluation project, we have acquired a database of 18 normal subjects. Each subject underwent a T1-MR (1.5T) SPGR 3D study. We have chosen arbitrarily a particular subject as the reference subject. For all methods, each subject (source image) is registered toward the reference subject (target image), so that all registration results may be compared in the same referential.

From these MR images, we have extracted anatomical features, that will be used to assess the quality of the registration processes. To be objective, the evaluation must rely on features that are independent of the similarity used to drive the registration process.

3.1 Tissue Classification

The most straightforward way to assess the quality of the registration is to evaluate how the tissues are deformed from one subject to the other. We extract grey matter and white matter from the MR volume using the method proposed in [6]. This algorithm consists first in a 3D texture analysis. A clustering technique gives a rough classification that is refined by a bayesian relaxation.

For each subject, we deform the grey and white matter classes toward the reference subject, using the deformation field and trilinear interpolation. The deformed classes are compared to the classes of the reference subject by computing overlapping measures [1]. For sake of concision, we only keep the total performance measure [1], and compute the mean and the variance of that measure over the database of 18 subjects.

3.2 L_{vv} Volume

We extract differential characteristics from the subjects with the L_{vv} operator, introduced by Florack *et al.* [5]. The sign of ML_{vv} has a very precise interpretation: it can be demonstrated that when limited to the cortical ROI the crest of a gyrus corresponds to a negative value of the ML_{vv} , while a deep fold like a sulcus corresponds to its positive part. Therefore, the sign of the mean curvature is sufficient to separate sulci from gyri [6].

For each subject, we deform the corresponding L_{vv} according to the results of a given registration method, using trilinear interpolation. We then compare this deformed L_{vv} with the L_{vv} volume of the reference subject, by computing a simple correlation. For each method, we compute the mean and the variance of that measure over the database of 18 subjects.

3.3 Extraction of Cortical Sulci

Cortical sulci are of great interest in the context of that paper, since they are relevant anatomical and functional landmarks. Due to the inter-individual cortical variability, the matching of sulci is crucial to evaluate different registration methods. Several methods have been developed to extract sulcal patterns from MR acquisitions. In this paper, we only describe rapidly the method we have used [6]. After a segmentation of cortical regions and cortical folds via differential operators, a compact and parametric description of a sulcus can be obtained by a medial surface representing the buried part of this sulcus. The method used here consists in modeling this surface by using an "active ribbon" which evolves, in the three-dimensional space, from a 1D curve to a 2D surface. The final position of the ribbon approximates the medial axe of the considered sulcus.

For each subject of the database, we extract 12 major sulci with the method described above. The sulci used for the evaluation project are central sulcus, precentral sulcus, postcentral sulcus, sylvian sulcus, superior frontal sulcus and superior temporal sulcus, for each hemisphere. For each subject, each sulcus is deformed toward the reference subject using the results of a given registration method. As the sulci are modeled by 3D B-splines, we deform each control point of the spline using trilinear interpolation, which naturally defines the deformed sulcus. A "distance" between the deformed sulci of each subject and the corresponding sulcus of the reference subject can be computed.

4 Results

4.1 Global Measures

Average Volume For each method, we deform each subject toward the reference subject, using the transformation and trilinear interpolation. It is finally possible to compute, for each method, a mean volume by averaging the 17 deformed subjects. A sagittal view of the average volumes are presented on figure 1, and can be compared to the corresponding view of the reference subject. Furthermore, we compute for each method the Mean Square Error (MSE) between the average volume and the reference volume, only for the voxels that belong to the brain of the reference subject (see table 1). The MSE is not a good measure to evaluate the quality of the registration of one subject, but is in that case a more relevant indication as we deal with average volumes.

It must be noted that the registration of the subject 9 has failed for the method A. Therefore, and for all the experiments, the subject 9 has been removed of the results of method A.

Overlapping of Grey and White Matter Tissues At that stage, the evaluation is not objective, as the MSE is more or less related to the similarity used to drive the registration processes, at least for the methods A, D, M and R. Therefore, we use the segmentation classes (grey matter and white matter) of each subject to evaluate how tissues overlap after registration, as described previously. Table 1 gives the mean and standard deviation of that measure over the database of subjects for each method.

The method M does not give very satisfactory results, whereas the methods D, P and R seem to give better and similar results. The method A seems to be slightly less

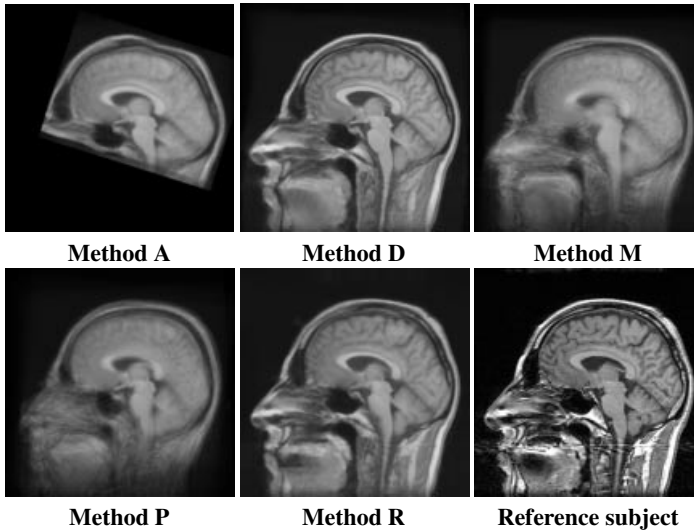


Fig. 1. For each method, the mean volume is obtained by averaging the 17 deformed subjects, and can be compared to the reference subject.

efficient, but we must keep in mind that the deformation field is computed at a 4 mm grid.

Correlation of L_{vv} The L_{vv} operator has been presented in the section 3.2, and provides information related to sulco-gyral patterns. For each method, the mean and standard deviation of the correlation coefficient (between deformed and original L_{vv}) are presented in table 1.

We first observe that the mean value of the correlation coefficient is quite low for all the registration methods. This might indicate that the matching of cortical features is not very good, but that point will be studied more extensively in section 4.2. The difference between the method M (mean value of 0.01) and other methods is significant. Method D seems to give a slightly better result with a mean correlation of 0.43.

4.2 Local Measures

Visualization of Deformed Sulci We have chosen first to visualize how each left central sulci of the 17 subjects deforms toward the left central sulcus of the reference subject (see figure 2, and associated caption for color code). For a perfect registration, the blue sulci should therefore be superimposed on the yellow sulcus.

It can be observed that the different registration methods seem to give a significant dispersion around the reference sulcus. The postcentral and precentral sulci of the reference subject (in red and green) give the order of magnitude of the dispersion, and indicate that in most cases, the position of the deformed sulci is misleading, with regards to the identification of sulci. If method M seems to give the highest variability, it is quite difficult to distinguish visually the performances of the methods A, D, P and R.

Method	Average volume	Method	Tissue	Mean	St. dev.	Method	Mean	St. dev.
A	987.9	A	grey	91.9	0.08	A	0.17	0.003
			white	89.6	0.07			
D	491.1	D	grey	95.8	0.04	D	0.43	0.005
			white	96.7	0.04			
M	1389.9	M	grey	88.8	0.13	M	0.01	0.001
			white	87.5	0.17			
P	1064.4	P	grey	93.5	0.06	P	0.16	0.003
			white	95.1	0.04			
R	385.6	R	grey	93.9	0.10	R	0.32	0.008
			white	95.0	0.14			

Table 1. Left: Mean Square Error (MSE) between the average volume and the reference subject. The error is computed only for the voxels that belong to the segmentation mask of the reference subject's brain. **Middle:** Overlap between tissues after registration, computed by the total performance measure. For each method, the mean and standard deviation of the measure is computed over the database of subjects. **Right:** Mean and standard deviation of the correlation coefficient between reference L_{vv} and deformed L_{vv} .

Numerical Evaluation Beyond visualization, numerical evaluation is needed. In that section, we investigate two measures: one for the global positioning of sulci, and one for shape similarity.

Euclidian distance between registered sulci To assess how sulci are matched, it is possible to compute an euclidian distance between a sulcus, deformed toward the reference subject, and the corresponding sulcus of the reference subject. As explained in section 3.3, sulci are modeled by B-splines, and may therefore be resampled identically. We associate the distance between sulci to the distance between control points.

To present a compact measure, the mean of the distance after registration are computed for all the subjects and all the sulci (we have 12 sulci extracted for each of the 18 subjects). These results are presented in table 2, and the distances are expressed in voxels (the resolution of the voxels is 0.93 mm). It can be immediately noticed that the results are not significantly different between rigid and non-rigid methods.

Statistical study of deformed shapes The distance between registered sulci is not a sufficient measure to characterize how sulci deform. We want to evaluate the similarity of deformed sulci in terms of shape, with the use of the Principal Component Analysis (PCA) [3].

For each method, we have a population of shapes that is composed by the corresponding sulci of the different subjects, deformed toward the reference subject by a given registration method. The purpose of the PCA is to analyze the variations of each shape with respect to the reference shape, by decomposition on the eigenvectors of the covariance matrix.

For sake of concision, we have chosen to consider only the trace of the covariance matrix. This measure reflects the entire variation of the population around the reference sulcus, along all the axes of the decomposition. Furthermore, the trace can be compared

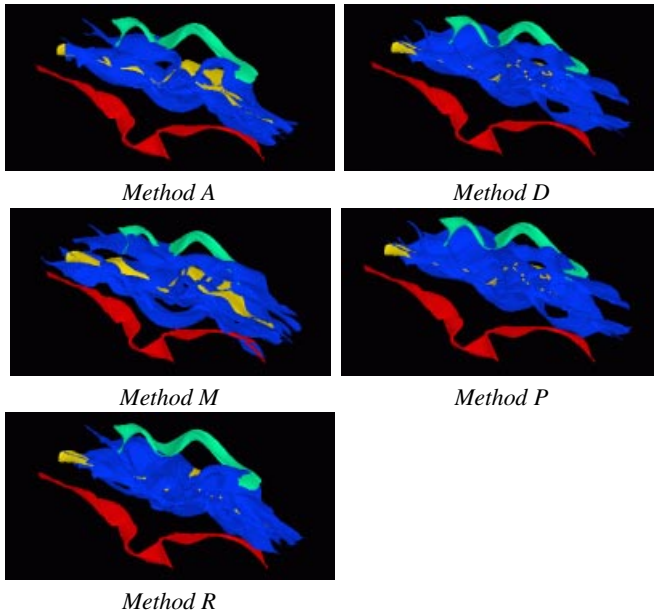


Fig. 2. Left central sulci (in blue) of the database deformed toward the reference subject. The deformed sulci should ideally be superimposed to the left central sulcus of the reference subject (in yellow). The left precentral sulcus (in red) and postcentral sulcus (in green) of the reference subject are also drawn.

since it is invariant when the axes of the decomposition change. These results are given on table 2. We notice that there is no significant difference between the performances of the different methods.

5 Conclusion

We have proposed in this paper an evaluation framework of methods that aims at registering brains of different subjects. Global and local measures of the registration have been designed to evaluate 5 registration methods on a database of 18 subjects. On the one hand, global measures show the efficiency of non linear methods, and indicate that the quality of the registration increases with the degrees of freedom of the estimated transformation. On the other hand, affine and non-linear methods give surprisingly similar results for local measures, which are based on the matching of major cortical sulci.

To explain these results, we must first keep in mind that the variability of cortical patterns between individuals is very high [11]. We are also tempted to put forward the anatomical "incorrectness" of transformations generated by "computer vision" methods. "Iconic" approaches, which tend to match voxels having the same luminance, fail to apprehend morphological differences between individuals because they use "low level" information. These results also stimulate the introduction of higher anatomical constraints, such as cortical constraints in the registration process.

Method	Average distance	Method	central	superior frontal	sylvian
A	9.9	A	547	736	1172
D	10.3	D	675	767	1046
M	11.5	M	621	622	1373
P	10.7	P	510	859	1233
R	10.8	R	735	741	1064

Table 2. Left: average distance between registered sulci and corresponding sulci of the reference subject, in voxels. The mean is computed for all the subjects and all the sulci. **Right:** for three different population of sulci, the variations of deformed sulci can be analyzed by principal component analysis. The trace of the covariance matrix, normalized by the number of subjects, traduces the entire variation of deformed sulci around the reference sulcus, in the shape space.

References

1. JH. Van Bemmelen, MA. Musen. *Handbook of medical informatics*. Springer, 1997.
2. L. Collins, A. Evans. Animal : validation and applications of nonlinear registration-based segmentation. *IJPRAI*, 8(11):1271–1294, 1997.
3. T. Cootes, C. Taylor, D. Hooper, J. Graham. Active shape models- their training and application. *CVIU*, 61(1):31–59, 1995.
4. A. Evans, L. Collins, B. Milner. A MRI-based stereotaxic atlas from 250 young normal subjects. *Soc. Neuroscience abstract*, 18:408, 1992.
5. L. Florack, B. Romeny, J. Koenderink, M. Viergever. Scale and the differential structure of images. *IVC*, 10:376–388, 1992.
6. G. Le Goualher, C. Barillot, and Y. Bizais. Modeling cortical sulci with active ribbons. *IJPRAI*, 8(11):1295–1315, 1997.
7. P. Hellier, C. Barillot, E. Mémin, and P. Pérez. Hierarchical estimation of a dense deformation field for 3D robust registration. In *IEEE TMI*, 20(5):388–402, 2001.
8. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens. Multimodality image registration by maximisation of mutual information. *IEEE TMI*, 16(2):187–198, 1997.
9. J. Maintz, MA. Viergever. – A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
10. J. Mazziotta, A. Toga, A. Evans, P. Fox, and J. Lancaster. A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage*, 2:89–101, 1995.
11. M. Ono, S. Kubik, C. Abernathy. – *Atlas of the cerebral sulci*. – Verlag, 1990.
12. J. Talairach, P. Tournoux. *Co-planar stereotaxic atlas of the human brain*. Georg Thieme Verlag, Stuttgart, 1988.
13. JP. Thirion. Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical Image Analysis*, 2(3):243–260, 1998.
14. P. Thompson, R. Woods, M. Mega, A. Toga. Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. *HBM*, 9:81–92, 2000.
15. P. Viola, W. Wells. Alignment by maximisation of mutual information. *IJCV*, 24(2):137–154, 1997.
16. J. West, J. Fitzpatrick, *et al.* Comparison and evaluation of retrospective intermodality brain image registration techniques. *JCAT*, 21(4):554–566, 1997.