Polyrigid and Polyaffine Transformations: a Novel Geometrical Tool to Deal with Non-Rigid Deformations - Application to the registration of histological slices

Vincent Arsigny, Xavier Pennec, Nicholas Ayache

Polyrigid and Polyaffine Transformations: a Novel Geometrical Tool to Deal with Non-Rigid Deformations. Application to the registration of histological slices.

Vincent Arsigny, Xavier Pennec, Nicholas Ayache

INRIA Sophia, Epidaure team, 2004 Route des Lucioles, F-06902 Sophia-Antipolis Cedex, France

Abstract

We describe in this paper a novel kind of geometrical transformations, named polyrigid and polyaffine. These transformations efficiently code for locally rigid or affine deformations with a small number of intuitive parameters. They can describe compactly large rigid or affine movements, unlike most free-form deformation classes. Very flexible, this tool can be readily adapted to a large variety of situations, simply by tuning the number of rigid or affine components and the number of parameters describing their regions of influence.

The displacement of each spatial position is defined by a continuous trajectory that follows a differential equation which averages the influence of each rigid or affine component. We show that the resulting transformations are diffeomorphisms, smooth with respect to their parameters. We devise a new and flexible numerical scheme to allow a trade-off between computational efficiency and closeness to the ideal diffeomorphism. Our algorithms are implemented within the Insight Toolkit, whose generic programming style offers rich facilities for prototyping. In this context, we derive an effective optimization strategy of the transformations which demonstrates that this new tool is highly suitable for inference.

The whole framework is exemplified successfully with the registration of histological slices. This choice is challenging, because these data often present locally rigid deformations added during their acquisition, and can also present a loss of matter, which makes their registration even more difficult.

Powerful and flexible, this new tool opens up large perspectives, in non-rigid 3D rigid registration as well as in shape statistics.

Key words: Parametric transformation, Diffeomorphism, Insight Toolkit, Non-rigid registration, Histological slices, Image registration.

1 Introduction

To motivate the introduction of a new type of geometrical transformation, let us focus on the registration of medical images, and more precisely on the types of geometrical transformations they are based on.

The registration of medical images is in general a difficult problem, and numerous methods and tools have been already devised to address this task (Maintz and Viergever, 1998). At the beginning of the spectrum, we have simple parametric transformations such as rigid or affine transformations, which have a very small number of degrees of freedom, and can be efficiently used for intra-patient registration. Other types of transformations, such as those parameterized via B-Splines (Rueckert et al., 1999), Thin-Plate-Splines (Bookstein, 1999), finite elements mechanical models (Ferrant et al., 1999) or more general deformable models can have an arbitrary number of degrees of freedom and be used for both inter-subject or intra-subject registration. At the end of the spectrum, deformation fields defining a displacement at every voxel exhibit the highest number of degrees of freedom (Thirion, 1998; Cachier, 2002; Cachier et al., 2003; Chefd'hotel et al., 2002; Hermosillo et al., 2002), and can be used for inter-subject registration.

Each of the above transformations has its particular domain of application. However, in the case of anatomical structures incorporating rigid elements (such as bone articulations, or structures which are subject to simple local deformations, like histological slices), we believe that none of them is fully appropriate. Rigid and affine transformations clearly don't have enough degrees of freedom. On the contrary, deformation fields have too many and thus can be easily misled by local minima of the similarity criterion. For the existing intermediate transformations, e.g. B-Splines (Rueckert et al., 1999), the degrees of freedom of the transformation are not really adapted since many control points are required to reconstruct several locally rigid behaviors, especially when rotations are substantial.

Our goal in this paper is to define new parametric transformations that exhibit a locally rigid or affine behavior, and that can be efficiently implemented. Also, a very desirable property is *invertibility*, which is not guaranteed in the approaches based on splines or other interpolation techniques, except in the case of the Geodesic Interpolating Splines (Camion and Younes, 2001), which are limited to the interpolation of a sparse set of displacements.

Email address: {Vincent.Arsigny, Xavier.Pennec, Nicholas.Ayache}@sophia.inria.fr (Vincent Arsigny, Xavier Pennec, Nicholas Ayache).

URL: http://www-sop.inria.fr/epidaure/index.php (Vincent Arsigny, Xavier Pennec, Nicholas Ayache).

An approach was proposed in Little et al. (1996) to smoothly interpolate a deformation outside independent rigidly moving regions. This computationally efficient approach is unfortunately "parameterized" by the motion *and the arbitrarily complex shape* of each rigidly-moving region. As a consequence, it is not straightforward to use this model for inference (i.e. non-rigid registration). Moreover, the invertibility of the interpolated transformation is not always ensured. This interpolation method is used in Pitiot et al. (2003), which deals with the registration of histological slices. This is a pivotal issue for the fusion of MR images and histological slices, which is a promising technique for building precise atlases of brain structures (Ourselin et al., 2001a; Bardinet et al., 2002).

Our idea is to use simple fuzzy regions defined by very few parameters: mainly the position of the center, a typical radius of influence and the associated rigid or affine transformation. We show in Section 2 that a simple average of the displacement induced by each region leads to invertibility problems. Thus, we develop an infinitesimal approach where the displacement is obtained by the integration of the average speed. To address the implementation efficiency, we investigate in section 3 several numerical schemes. The result is a new family of invertible and fully parametric transformations that we called *polyrigid and polyaffine transformations*. The Insight Toolkit (ITK) is a very attractive framework for the implementation of these new transformations, since it provides a base class for all parametric transformations and powerful registration tools that greatly facilitate the rapid development of new algorithms. We describe also in Section 3 how Polyrigid Transformations are implemented within this framework. One can find the code relative to the experiments conducted in this paper on the Internet [1] . We show in Sec. 4 that this new general tool is well-suited for the non-rigid registration of articulated-like object. This is exemplified on 2D histological slices. In Section 5, we also present preliminary results that show how polyrigid transformations can be refined to describe precisely regions of influence of a complex shape.

2 Theory of Polyrigid and Polyaffine Transformations

2.1 Regions of Influence and Interpolation of Sparse Data

2.1.1 Simple Parameterization of Regions of Influence

In order to model transformations having several distinct rigid behaviors in different regions, it is necessary to define how each component of the global

[1] *ftp://ftp-sop.inria.fr/epidaure/Softs/Arsigny/ITK_PolyTransfoSRC*

transformation is anchored geometrically. One could of course choose to have regions of influence of arbitrary shape, like in Little et al. (1996), but this is not convenient for inference. Having a reduced number of parameters describing the shape and extent of each region of influence allows for simple optimization of these parameters, which is a highly desirable feature for registration purposes.

We propose here a Gaussian model for regions of influence: to each region we have an anchor point $a \in \mathbb{R}^n$, and in addition we also have two other parameters, a typical distance σ and a parameter p such that the influence of the i-th component is described by a "weight" $w_i(x) = p_i\, G_{(a_i,\sigma_i)}(x)$ where $G_{(a_i,\sigma_i)}$ is the Gaussian of mean a_i and of standard deviation σ_i. Thus, instead of using regions in which the transformation is purely rigid like in Little et al. (1996), we propose "fuzzy" regions, which makes the transitions or interpolations between the regions straightforward to handle.

In order to obtain a global transformation from several weighted components, the classical way of mixing each local behavior is given in Sheppard (1968), which simply amounts to averaging the displacements according to the weights:

$$T(x) = \frac{\sum_i w_i(x)\, T_i(x)}{\sum_i w_i(x)}. \tag{1}$$

Here, the transformations $(T_i)_{i\in 1\cdots N}$ are rigid transformations. They are parameterized by the rotation matrixes (R_i) and the translations (t_i). Their action on a point is given by:

$$\forall x \in \mathbb{R}^n,\ T_i(x) = R_i\, x + t_i.$$

2.1.2 Weaknesses of the Classical Averaging

The transformation obtained via (1) is smooth, both with respect to spatial coordinates and its parameters. Nonetheless, it has several major drawbacks:

- Its invertibility is not guaranteed, and indeed will not be assured in many cases, for example if the displacements are large.
- In the favorable case where the inverse exists, it has in general no simple form and has to be estimated by an *ad hoc* technique, for instance using a general deformation field, which is iteratively optimized to obtain the inverse
- Is the behavior of this direct averaging procedure really qualitatively satisfactory? In Fig. 1, an example shows that in the case of a mixture of rotations, points do not in general turn around the centers of the rotations. On the contrary the approach proposed here has this property.

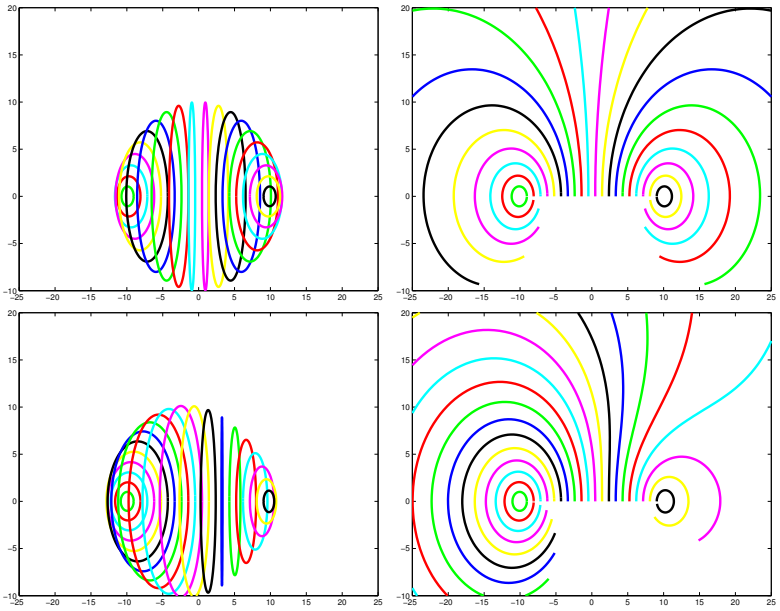

Fig. 1. **Simple averaging (left) versus proposed approach (right) .** Here, the polyrigid transformation has two rotation components, which have exactly opposite angles. We consider in this figure the various trajectories of points originally in the segment joining the two centers. These trajectories are constituted by all the final positions of the initial points as we progressively increase the angle of rotation from 0 to 2π radians. On top, the two relative weights p_1 and p_2 are equal whereas on the bottom that of the left component is substantially higher than the other, hence the greater influence of the transformation anchored in the left. The form of trajectories show that points moving under the action of a polyrigid transformation do turn around the centers of rotations of the transformation. This property is not verified in the case of the classical averaging.

These reasons have led us to develop a new kind of averaging procedure tackling the above-mentioned problems.

2.2 A Framework with ODEs

2.2.1 Invertibility and ODEs

The challenge facing us at this point is the following: how to mix several transformations according to some weight functions in an invertible way? A classical way of obtaining invertible and smooth transformations is to use ordinary differential equations (ODEs). We refer the reader to Tenenbaum and Pollard (1985) for the following classical results on ODEs. A particle governed by an ODE follows an equation of the form:

$$\dot{x}(s) = V(x, s).$$

The *flow* associated to this ODE is the function that maps to a given starting position x_0 the position $\Phi(x_0, s)$ reached at time s by the particle following the evolution prescribed by the ODE. In other words, the trajectory $x(s) = \Phi(x_0, s)$ is a solution of $\dot{x}(s) = V(x, s)$.

If V is smooth (for instance $\mathcal{C}^1$) with respect to x (spatial coordinates) and s (time), *and if the solution $x(s)$ is defined for all time*, then the *flow* $\Phi(x, s)$ associated to the ODE defines a family of diffeomorphisms. More precisely, for each $s \in \mathbb{R}$, we have that $x \mapsto \Phi(x, s)$ is a diffeomorphism of $\mathbb{R}^n$ in $\mathbb{R}^n$. Our approach is based on this key result.

2.2.2 The Case of Rigid Transformations

From the classical results of linear algebra, it is obvious that a rigid transformation is invertible, and its inverse is simply obtained by inverting the rotation part and adapting the translation component in the appropriate way. But another viewpoint can be used to prove the invertibility, using ODEs. More precisely, we can associate to a rigid transformation the following ODE, where the nature of the A_i matrix is explained just below:

$$\dot{x}(s) = V_i(x, s) = t_i + A_i\,(x - s\,t_i) \text{ for } s \in [0, 1]. \tag{2}$$

This is obtained by differentiating the trajectory equation $x(s) = s\ t_i + \exp(sA_i)\ x(0)$. At time 0, we start with the initial position and the image for the rigid transformation is obtained at time 1. Since V_i is smooth and trajectories are defined for all time, the above-mentioned result applies.

In Eq. (2), we denote by A_i one of the logarithms of the rotation matrix R_i, which verifies the equality: $\exp(A_i) = R_i$ where exp is the matrix exponential. Since R_i is a rotation, it always has a real logarithm, which is a skew matrix. For example, let in 3D $r = (r_x, r_y, r_z)^T$ be a rotation vector associated to a rotation R. We can then define a skew matrix A associated to r that is a logarithm of R with the relation:

$$A = \begin{pmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{pmatrix}.$$

2.2.3 A Continuous Averaging Procedure with ODEs

In order to ensure the invertibility of our averaged transformation, let us define a new ODE. The idea is simply to average according to weights the *speed*

vectors associated to each component, instead of averaging the final results:

$$\dot{x}(s) = V(x,s) = \frac{\sum_i w_i(x)\, V_i(x,s)}{\sum_i w_i(x)} \tag{3}$$

Ideally, we would like to define our averaged transformation as $T(x) = \Phi(x,1)$, where Φ is the flow associated to the ODE (3).

This means that each component will influence the motion of a particle accordingly with the weights modeling its influence in space. The result obtained at time 1 is the image of initial position x under the action of the average transformation.

2.3 Theoretical Properties of Polyrigid Transformations

2.3.1 Life-Span of a Solution to an ODE

As mentioned before, in order to define our average transformation, it is necessary to prove that the position at time 1 exists, whatever the initial position may be. For an arbitrary ODE, the existence is not always ensured, however smooth the speed function V may be. Consider for instance, the 1D evolution

$$\dot{y}(s) = V(y) = y^2.$$

Its solution with an initial position y_0 is $y(s) = \frac{1}{1/y_0 - s}$. Thus, we see that for $1/y_0 > 0$, the life-span of the solution only extends between $-\infty$ and $1/y_0$, and if $1/y_0 < 1$, then the position at time 1 is absolutely undefined, the particle having gone to infinity before that!

2.3.2 Existence and Invertibility of Polyrigid Transformations

Since in Eq. (3) $V(x,s)$ is $\mathcal{C}^\infty$ with respect to spatial position and time, it only remains to be proved that the evolution does not lead to explosion towards infinity before time 1.

Theorem 1. All solutions of Eq. (3) have an infinite life-span, i.e. they are defined for all time, whatever the rigid transformations may be. The polyrigid transformations defined via $T(x) = \Phi(x,1)$ are thus well-defined and diffeormorphic.

Proof. There exists three positive constants C_1, C_2 and C_3 such that: $\|V(x,s)\|_2 \leq C_1 + C_2|s| + C_3\|x\|$. For instance, take $C_1 = \max_i \|t_i\|_2$ and $C_2 = \max_i \|A_i t_i\|_2$ and $C_3 = \max_i \|A_i\|_2$ where $\|A_i\|_2$ refers to the Frobenius norm of matrix A_i, equal here to the L^2 norm of the associated rotation vector.

This yields via a classical bounding that $\forall s,\ \|\Phi(x,s)\| \leq e^{C_3|s|}(\|\Phi(x,0)\| + (1-e^{-C_3|s|})(C_1/C_3 + |s|C_2/C_3))$, which suffices to prove the result because it shows that the position of the particle evolving with Eq. (3) is contained within a sphere whose radius grows exponentially this time. Thus, for any finite time, $\Phi(x,s)$ is at a finite distance of the initial solution x. According to the previous results from the classical ODE theory (Section 2.2.1), we have that $T(x) = \Phi(x,s)$ is a diffeomorphism.

A simple inverse. The inverse of the transformation is obtained here in a simple fashion: it suffices to go back in time! The skew matrix is changed into its opposite, the translation also and s becomes $1-s$. The inverse transformation thus takes here a simple form.

2.3.3 Differentiability with Respect to the Parameters

We have just seen that any given system of rigid transformations can be averaged so as to yield a diffeomorphism. But, what smoothness can be guaranteed with respect to the parameters? Differentiability is crucial so as to enable simple optimization of the transformation in a registration framework. We have the following result:

Theorem 2. Polyrigid transformation are $\mathcal{C}^\infty$ with respect to all parameters.

Proof. This comes from the differentiability of the flow of an ODE. Indeed, let us define the new ODE $\dot{z}(s) = W(z,s)$ where $z = (x,p)$, x being the spatial coordinates of a particle and p the parameters of the polyrigid transformation written in a vectorial fashion, and where the speed vector $W(z,s) = (V(x,s),0)$. Thus, x evolves according to (3) and that p does not change as time goes by. W is $\mathcal{C}^\infty$ and the solutions are defined for all time since those of (3) are. This implies the differentiability of the flow associated to this ODE, which is exactly the differentiability of the polyrigid transformation with respect to its parameters.

2.4 Extension to Polyaffine Transformations

2.4.1 A Simple Extension via the Real Logarithm

One can wonder to what extent it is possible to use the framework presented above to work with locally affine transformations. This can be done in a direct way if each affine transformation (M_i, t_i) has a linear part M_i that admits a *real logarithm*, i.e., if there exists a $n \times n$ real square matrix A_i such that $\exp(A_i) = M_i$. Then, we can adopt all coordinates of A_i as new scalar parameters to work with, and all the results of this section hold for this other type

of transformation, that we could call polyaffine.

Unfortunately, not all real invertible matrices M_i have a real logarithm. Even among real matrices with a positive determinant, this is not true. This is unsatisfactory, because compositions of a dilatation and a rotation are deformations that are essential to the affine generalization of polyrigid transformations.

2.4.2 The general extension

In order to define polyaffine transformations, what we basically need is simply a smooth trajectory from the identity to any given affine transformation. By differentiating with respect to time we want to obtain a simple ODE. To define such an evolution, we have the following general result: any element of a real connected Lie Groups is equal to the product of *two* exponentials (Wüstner, 2003). Indeed, in the linear part of affine transformations, the singular value decomposition yields that $M_i = \exp(A_i)\,\exp(S_i)$ with A_i and S_i respectively skew and symmetric matrices. An equivalent of Eq. (2) for polyaffine transformations is thus:

$$\dot{x}(s) = t_i + (A_i + \exp(sA_i)\,S_i\,\exp(-sA_i))\,(x - s\,t_i). \tag{4}$$

All results mentioned above still hold in this case and hence we can define general polyaffine transformation, smooth both w.r.t. spatial coordinates and parameters. But other parameterization could be chosen: we can also write $M_i = \exp(\tilde{S}_i)\,\exp(A_i)$ by regrouping the factors of the SVD differently, where $\tilde{S}_i \neq S_i$ in general. Several extensions are possible and will be investigated in future work.

2.5 Summary of the properties of Polyrigid Transformations

In this section, we defined a new class of transformations, modeling a mixture of rigid transformations, whose influence is geometrically anchored in a simple way. These transformations are diffeomorphisms and smooth with respect to all of their parameters. The following tables summarize the various parameters of the transformations (Table 1), and the number of scalar parameters obtained in 2 dimensions or 3 dimensions (Table 2), where a comparison is be made with B-Splines.

Region parameters	Deformation parameters
Anchor points: (a_i)	Rotation vectors: (r_i)
Standard deviations: (σ_i)	Translation vectors: (t_i)
Relative weights: (p_i)	

Table 1
The two types of parameters of polyrigid transformations.

Num. regions	DOF in 2D	Equivalent B-Spline	DOF in 3D	Equivalent B-Spline
2	13	3 control points	21	3 c.p.
3	20	5 c.p.	32	5 c.p.
4	27	6 c.p.	43	7 c.p.
N	$7N-1$	$\frac{7N-1}{4}$ c.p.	$11N-1$	$\frac{11N-1}{6}$ c.p.

Table 2
Number of parameters (DOF) of polyrigids with different number of components, and comparison with the number of control points of the equivalent spline (in term of degrees of freedom).

3 Implementation of Polyrigid Transformations

3.1 Discretization Schemes

Since in the general case there does not exist a closed form for the position of a point moving under the action of a polyrigid transformation, it is a necessity to resort to a numerical scheme to integrate the ODE defining the transformation. In other words, the trajectory of a point moving via (3) has to be sampled: a number of intermediate points N and a rule for obtaining the successive positions $(x_i)_{i\in 0\cdots N}$ have to be chosen, so that the curve defined by the points converges toward the real continuous curve given by the ODE.

In our domain of application, i.e. medical imaging, we have an additional constraint, due to the volume of data much must be processed in common applications. Thus, the numerical scheme should be as computationally inexpensive as possible. This is all the more true here that we will use in Section 4 the first and second derivatives of the numerical scheme to optimize the transformation parameters during registration, which forbids the use of classical schemes such as Runge-Kutta's.

3.1.1 The Consistent First Order Scheme

The consistence of a numerical scheme is a crucial notion. It is a condition that must be verified to ensure the convergence towards the continuous solution when the time step goes to zero. It simply means that when we take the Taylor expansion of the solution of the ODE with respect to time around zero, a numerical scheme must have the same expansion up to a certain order. We say also that a scheme is of a certain order when the coefficients of its Taylor series vanish after that order.

The consistent first order scheme is simply given in the following way: we define the operators $T_1^{1/N}$ and $T_1^{k/N}$ by:

$$\begin{cases} T_1^{1/N}(x,s) = x + \frac{1}{N}V(x,s). \\ T_1^{k/N}(x) = \underbrace{T_1^{1/N}(.,(k-1)/N) \circ \cdots \circ T_1^{1/N}}_{k \text{ compositions}}(x,0). \end{cases} \tag{5}$$

The points (x_i) are obtained recursively using:

$$\begin{cases} x_0 = x. \\ \text{for } 1 \leq k \leq N: \ x_k = T_1^{1/N}(x_{k-1},(k-1)/N) = T_1^{k/N}(x_0). \end{cases} \tag{6}$$

This simply means that starting at x_0, we jump from x_{k-1} to x_k by adding $1/N$ times the speed vector $V(x_{k-1},(k-1)/N)$.

3.1.2 An Efficient Second Order Scheme

The scheme described above is not really satisfactory. In the case of a single rigid component, the approximation makes points move along a diverging spiral instead of a circle (if the transformation is a rotation). This is regrettable, and a simple way of suppressing this approximation is to use the following second-order scheme using new operators $T_2^{1/N}$ and $T_2^{k/N}$:

$$\begin{cases} T_2^{1/N}(x,s) = x + \frac{\sum_i w_i(x)\left(\frac{1}{N}t_i + (\exp(A_i/N) - Id)(x - st_i)\right)}{\sum_i w_i(x)}. \\ T_2^{k/N}(x) = \underbrace{T_2^{1/N}(.,(k-1)/N) \circ \cdots \circ T_2^{1/N}}_{k \text{ compositions}}(x,0). \end{cases} \tag{7}$$

Instead of averaging the speed vectors of each component, we average instead the displacements that would be observed if each component was acting alone during a small interval of time of length $1/N$. This scheme is first-order consistent, but not second-order consistent (it captures only part of the second-order terms). But it is *exact* in the case of a single component, and its convergence is

much faster than the first one as shown in figure 2. Furthermore, the diverging spiral phenomenon observed for the first scheme disappears.

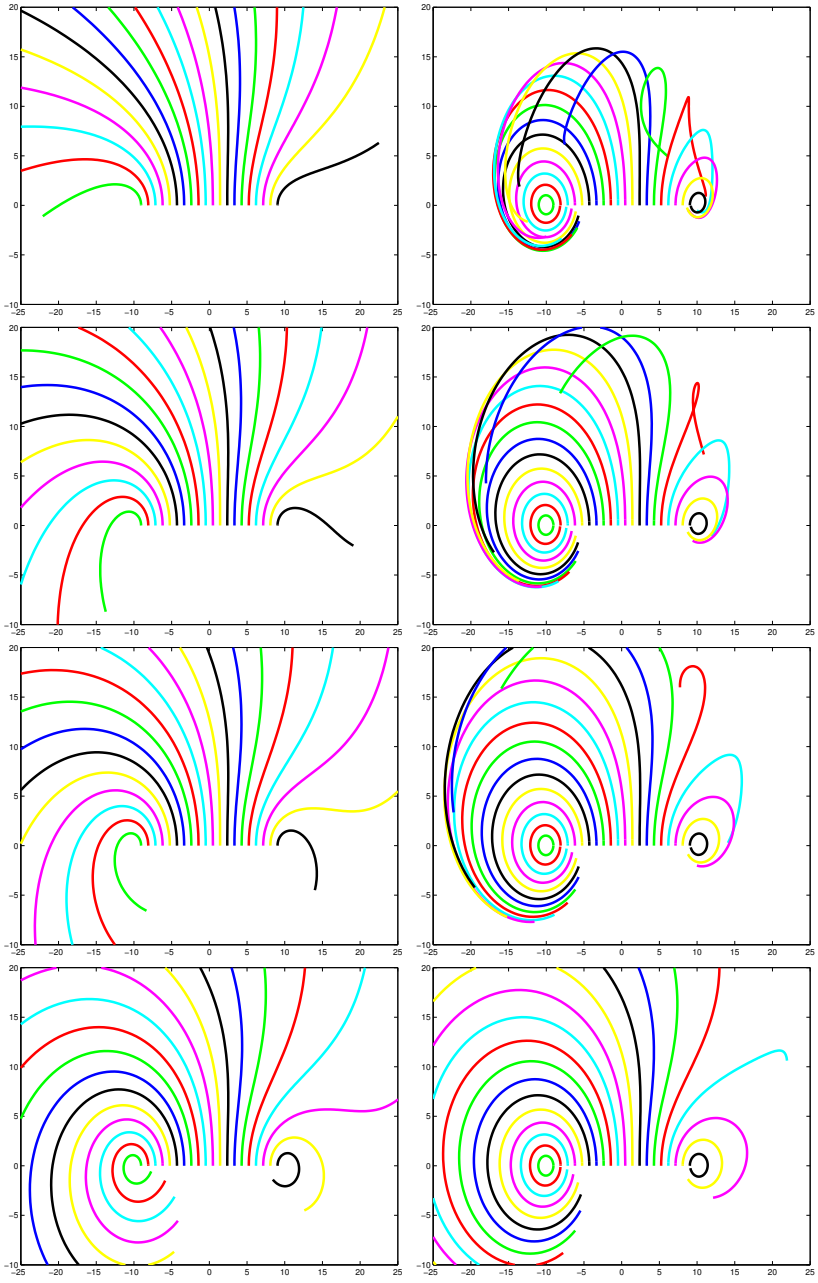

Fig. 2. **First scheme (on the left) versus second scheme (on the right).** From top to bottom: discretization levels of 3, 5, 7 and 20. As in figure 1, various trajectories are displayed, these trajectories being obtained when the two opposite rotations see their angle increase progressively between 0 and 2π. Here, the rotation on the left has a larger relative weight than that on the right, which lessens the influence of the latter.

3.1.3 Derivatives of the Transformation

Computing the derivatives of the transformation with respect to its parameters is necessary to use a gradient descent approach. Let us consider for instance a simple registration strategy, where we want to register two images I and J with the sum of square differences (SSD) criterion. This does not imply that our approach is restricted to that particular case: one could obviously compute the derivatives for other criteria. We take two images, J and I, and we want to register J onto I using the inverse of a polyrigid transformation T_p. p are the parameters of the transformation. In this case, the criterion to be minimized is:

$$S(I, J \circ T_p) = \int_\Omega \|I(x) - J \circ T_p(x)\|^2 \, dx.$$

The gradient of S with respect to p is the following:

$$\frac{\partial S}{\partial p}(I, J \circ T_p) = 2 \int_\Omega (J \circ T_p(x) - I(x)).(\nabla J \circ T_p)(x).\frac{\partial T_p}{\partial p}(x) \, dx.$$

In the last equation, the symbol "." denotes the matrix product. In order to compute the derivatives of the transformation with respect to the parameters, we simply computed the derivatives of each of the schemes. This is done again with a recursive formulation:

$$\frac{\partial T_p^{\frac{k}{N}}(x)}{\partial p} = \frac{\partial T_p^{\frac{1}{N}}\left(\cdot, \frac{k-1}{N}\right)}{\partial p}\left(T_p^{\frac{k-1}{N}}(x)\right) + \frac{\partial T_p^{\frac{1}{N}}\left(\cdot, \frac{k-1}{N}\right)}{\partial x}\left(T_p^{\frac{k-1}{N}}(x)\right).\frac{\partial T_p^{\frac{k-1}{N}}(x)}{\partial p}.$$

For a first-order gradient descent, only the above gradient is necessary. But for a second-order gradient descent, we will also need the second-order derivative:

$$\begin{aligned}\frac{\partial^2 S}{\partial p^2}(I, J \circ T_p) = 2\Big\{ & \frac{\partial T_p}{\partial p}(x)^T.(\nabla J \circ T_p)(x)^T.(\nabla J \circ T_p)(x).\frac{\partial T_p}{\partial p}(x) \\ & + (J \circ T_p(x) - I(x))\frac{\partial T_p}{\partial p}(x)^T.(\frac{\partial^2 J}{\partial x^2} \circ T_p(x)).\frac{\partial T_p}{\partial p}(x) \\ & + (J \circ T_p(x) - I(x))\frac{\partial T_p}{\partial p}(x).\frac{\partial^2 T_p}{\partial p^2(x)}\Big\}.\end{aligned} \tag{8}$$

A useful approximation is obtained by keeping only the first term of this equation. It has the nice property of being symmetric positive, and is a good approximation of the Hessian as long as that the difference of intensities $(J \circ T_p(x) - I(x))$ is small. Therefore, the more we will be close to a "good" solution, the more valid this approximation is. For detailed formulas, we refer the reader to Appendix A.

In order implement these new transformations, we chose to use the framework of the Insight Toolkit [2], which is a rich and rapidly developing set of tools dedicated to the segmentation and registration of medical images. Thanks to the generic nature of its programming style, it was the ideal choice to develop our new approach quickly.

The polyrigid transformations were implemented as a new transformation class: `PolyRigidTransform<TScalarType,Dimension,Dimension>`. It is templated by the dimension of the space, and thus can be used in both 2D and 3D applications. Its testing and the development of related registration algorithms were greatly facilitated by ITK, since it provides many tools that can be applied to *any* ITK transformation.

Figure 3 shows the registration framework chosen by ITK. The experiments presented in next section are carried out using our new class of transformation, the SSD similarity criterion (called here a "metric"), with a bilinear interpolation.

For the first-order gradient descent, we used the already implemented ITK optimizer `itk::RegularStepGradientDescentOptimizer`, in which the step of the gradient is reduced if the change of direction is too abrupt. This prevents the algorithm from going systematically too far in the direction of the gradient.

For the second-order gradient descent, we have implemented our own optimizer. This enabled us to adapt completely the optimization to the registration strategy studied in Section 4. Fig. 4 presents the new ITK classes designed to this effect. In order to take into account the information given by an approximation of the Hessian, we chose to modify the `itk::ImageToImageMetric` class, which provides access only to the first derivative of the similarity measure. Other classes were also modified, in order to perform a registration procedure making use of the Hessian, which is handled by the new class `ImageRegistrationMethodWithHessian`.

[2] http://www.itk.org

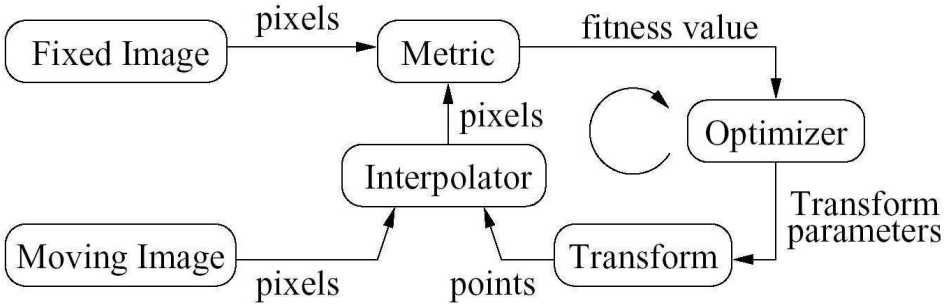

Fig. 3. ITK's registration framework.

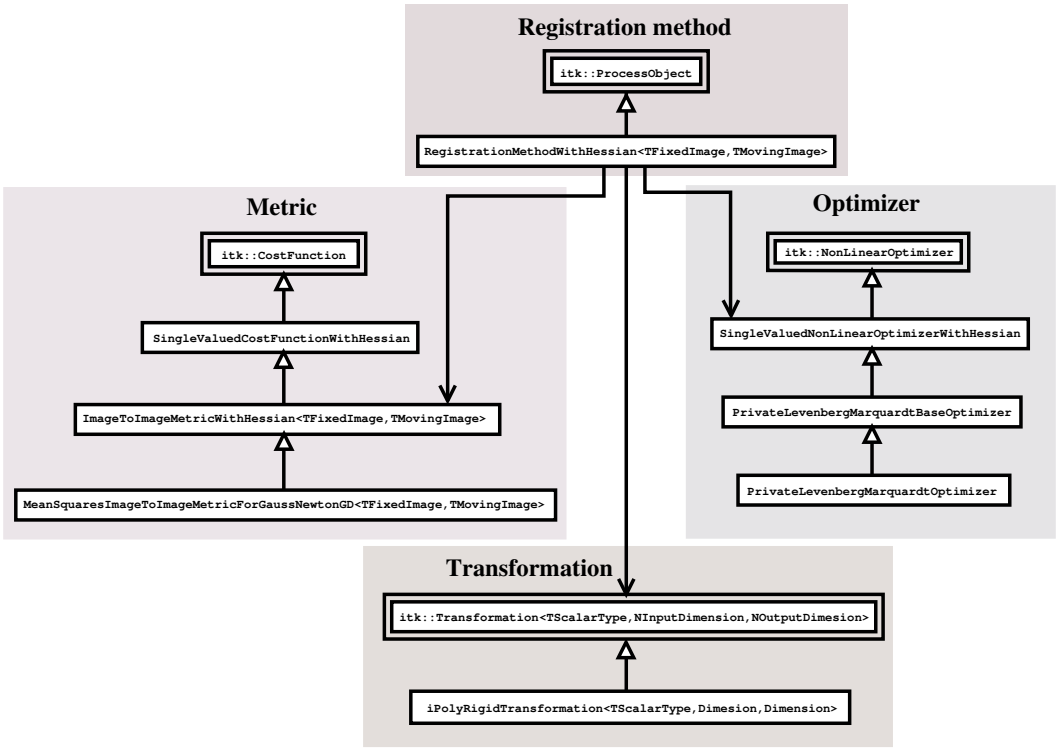

Fig. 4. **Implemented classes (boxes with a single rectangle) and their relations to each other and existing ITK classes (boxes with two nested rectangles)**. A hollow triangle at the end of an arrow stands for inheritance and and simple lines for dependence (conventions of the UML 1.3 standard). The classes belong to four different "families": that of the registration method, metric, optimizer and transformation. For more clarity, these groups have been put into boxes.

4 Registration of Histological Slices

4.1 Object of the Study and Experimental Setup

In order to demonstrate the feasibility and power of polyrigid transformations for registration purposes, we present in this section some preliminary results on

the registration of histological slices (Fig. 5). These images are acquired in such a way that locally rigid or even affine deformations are frequently introduced locally during the acquisition process. E.g., a gyrus has been rotated in the top left corner in Fig. 5. The aim of this study is to show that simple polyrigid transformations can substantially and naturally reduce the impact of such non-linear deformations, while preserving the anatomical differences, i.e. without introducing unrealistic deformations.

The slices studied in this paper have been kindly provided by P. Thompson and A. Pitiot from the LONI (UCLA), and are consecutive myelin-stained histological sections (or autoradiographs). Stemming from a human brain, it is during three steps that the artificial deformations of a locally rigid nature are introduced. These steps are the cutting process, the successive chemical treatments, and the glass mounting step. The dimensions of the slices are 226 by 384 pixels. The registration of these slices has an additional difficulty: the absence of matter in the lower-left-hand corner of the second slice. Many non-rigid registration algorithms are misled by such a defect because they will try to correct it, and in so doing they introduce irrelevant artificial deformations. During the acquisition process, the calibration of the optical setup remained unchanged. Therefore, the assumption that the various structures present in the images have the same grey level is valid, and we can safely use the sum of square differences criterion.

In figure 5, we see the results obtained with classical robust rigid and affine registration procedures Ourselin et al. (2001b). These methods are not able to register properly the rotated gyrus and at the same time all other gyri. This defect is due to the lack of degrees of freedom in these linear transformations. In the affine case, it is also due to the fact that the extra degrees of freedom, modeling dilatations and shearing, are not used to model the actual deformations appearing in the image. This suggests to use transformations with more degrees of freedom, and if possible, degrees of freedom that are adapted to the real deformations observed. This is precisely what polyrigid transformations are aiming at for this application.

During the experiments, the initialization used is the following:

- All rigid components are initially set to the identity.
- Anchor points are sampled on a regular grid, except in the first experiment of Section 4.2, where a manual initialization is done.
- Initial relative weights are all equal.
- (σ_i) are initialized at a high value, here 40, so that the influence of all regions extends in a good half of the images.

Four rigid components are used in the experiments. This number is a good compromise between the necessity of having enough degrees of freedom to reg-

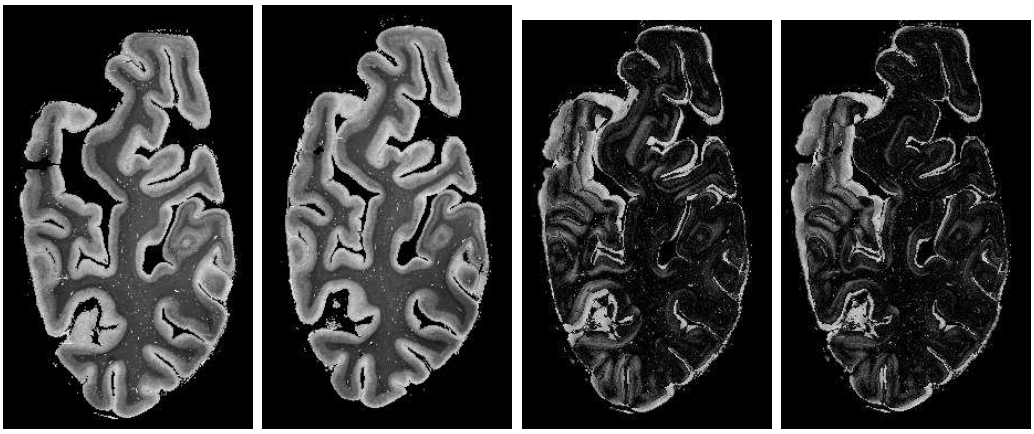

Fig. 5. **Histological slices (on the left) and Images of absolute differences for affine (third starting from the left) and rigid (on the right) registrations**. The whiter the grey level, the worse the registration is locally. We see in both cases that in many places, the edges of the gyri have not been registered precisely, because of the influence of the rotated gyrus in the top left corner and also because some other (smaller) non-linear deformations have taken place. We see also on the left that in order to register better the rotated gyrus, the affine registration gives poorer results for many edges than the rigid registration. Indeed, this better registration of the gyrus has been obtained at the cost of a dilatation of the slice, which in this situation is not appropriate.

ister correctly the slices and the obvious risk of introducing too many degrees of freedom, which results in large unrealistic deformations. This is precisely what occurs when more components are used. All in all, these four components are parameterized by 27 scalar parameters, which is the equivalent of only 6 control points for the B-splines (24 scalar parameters).

The second numerical scheme is used here, since it outperforms the other. The level of discretization is chosen very low, i.e. almost all results are obtained using no intermediary point between the starting position and the final position of a point. The deformed grids of Figures 7, 10 and 11 show that the obtained transformations are invertible (no self-intersection). However, if we had to use the inverse transformation, a few intermediate points would probably be necessary to ensure a good accuracy. In fact, determining the optimal number of points to sample the trajectories is still an open question that needs to be addressed, as in many diffeomorphic registration algorithms. In our case, we observed that increasing the number of discretization points used leads to very similar results, which shows that such a precision was not necessary here. But for other applications, if discontinuities appear or if it is necessary to use also the inverse transformation, then a finer discretization is of course essential.

4.2 Limitations of the First-Order Gradient Descent

A simple way to minimize the similarity criterion between the images is to use a first-order gradient descent, i.e. to make the parameters evolve in the steepest direction of descent, which is given by the gradient. Unfortunately, this approach cannot be directly used for our model. The partial derivatives in the gradient show differences of several orders of magnitudes! Qualitatively we have $\|\partial T_p/\partial r_i\| \gg \|\partial T_p/\partial t_i\| \gg$ other derivatives. This implies that the classical gradient descent will make rotations evolve enormously, the translations a little, and the other parameters almost not.

For rotations and translations, the difference of magnitude of their respective partial derivative can be intuitively understood in the following way: for a small variation of the angle of rotation, points far away from the center of rotation will move proportionally to their distance to the center of rotation. In other words, the further away from the center, the higher will be the variation in position, and a small variation can result in a large one at a distance. On the other hand, a variation in translation will affect all the points uniformly, and a small variation always yields a small modification in position. Therefore, we tend to have large partial derivative with respect to rotations as compared to partial derivatives with respect to translations. This difficulty is often encountered in situations where parameters of different natures are to be optimized simultaneously.

A simple remedy is to renormalize the amplitudes of the partial derivatives. Typically, dividing the amplitude of the rotation partial derivative by a factor 1000 is needed to obtain the optimization at least of both rotations and translations. Fig. 6 shows the behavior of the registration as the scaling evolves.

As we see in the deformed grid of Fig. 7, the final transformation is notably non-linear. But the anchor points have not moved from their initial position, which does not allow for an accurate registration in the upper left-hand corner. We can see in Fig. 8 that the edges were much better registered than with using a robust rigid transformation. But the incapacity to optimize the regions of influence thwarted the better registration of the upper-left-hand corner.

Of course, one could think of estimating a relevant renormalization for each type of parameter. This could be done by computing some kind of average amplitude for each partial derivative, and them dividing the derivatives by that value so as to obtain values of approximately the same amplitude. But this renormalization would have to be carried out for each pair of images to be registered. It would surely not be efficient for all iterations, and it is not so clear why all partial derivatives should have approximately the same amplitude. In the case of a pure translation, forcing the rotation vectors to evolve would not

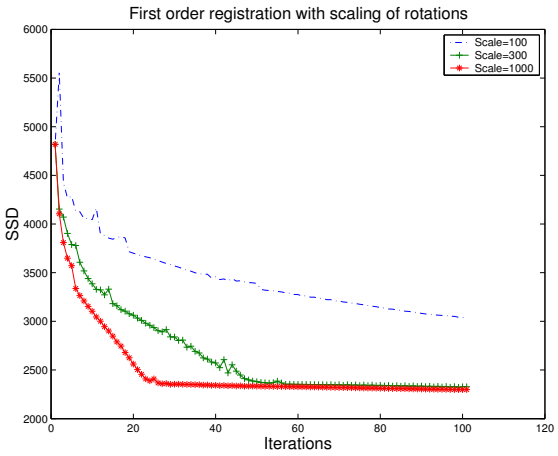

Fig. 6. **SSD criterion evolution for a polyrigid registration with a simple first order gradient descent.** The only modification to the gradient was the rescaling of the rotation partial derivative, which is much larger in magnitude than the others. The figure shows the SSD evolution during registration for three values of the rotation scaling: 100, 300 and 1000. Thus, we see that an important rescaling (at least of a factor 300) is necessary to improve the registration process, which is otherwise inefficient. The registration results only in the optimization of the rotations, the other parameters hardly evolving during the registration procedure.

be convenient! This calls for some type of adaptive renormalizing method.

4.3 Registration Results using a Levenberg-Marquardt Algorithm

To renormalize the partial derivatives in an adaptive way, a simple idea is to perform a second-order gradient descent scheme. The renormalization is handled by multiplying the gradient by the inverse of a matrix reflecting the second variations of the criterion. Here, this matrix is the approximation of the Hessian described in Section 3. The computation of this positive matrix term (in the sense of quadratic forms) can be done only at the expense of a very little cost, since it only requires the knowledge of the transformation's gradient and of the images intensities.

In order to perform an efficient 2nd-order gradient descent, the Levenberg-Marquardt algorithm (LM) was used (see Bazaraa et al. (1993), pages 312-314). At each iteration, a trust indicator is updated, which tunes the gradient descent between a simple first-order gradient descent and a quasi Gauss-Newton descent based on the truncated Hessian. This way, we obtain naturally a renormalization of the various parameters and also a faster convergence, especially when we are close to the minimum.

Figures 8 and 9 show that the Levenberg-Marquardt performs much better than a first-order descent, both quantitatively and qualitatively. Three major

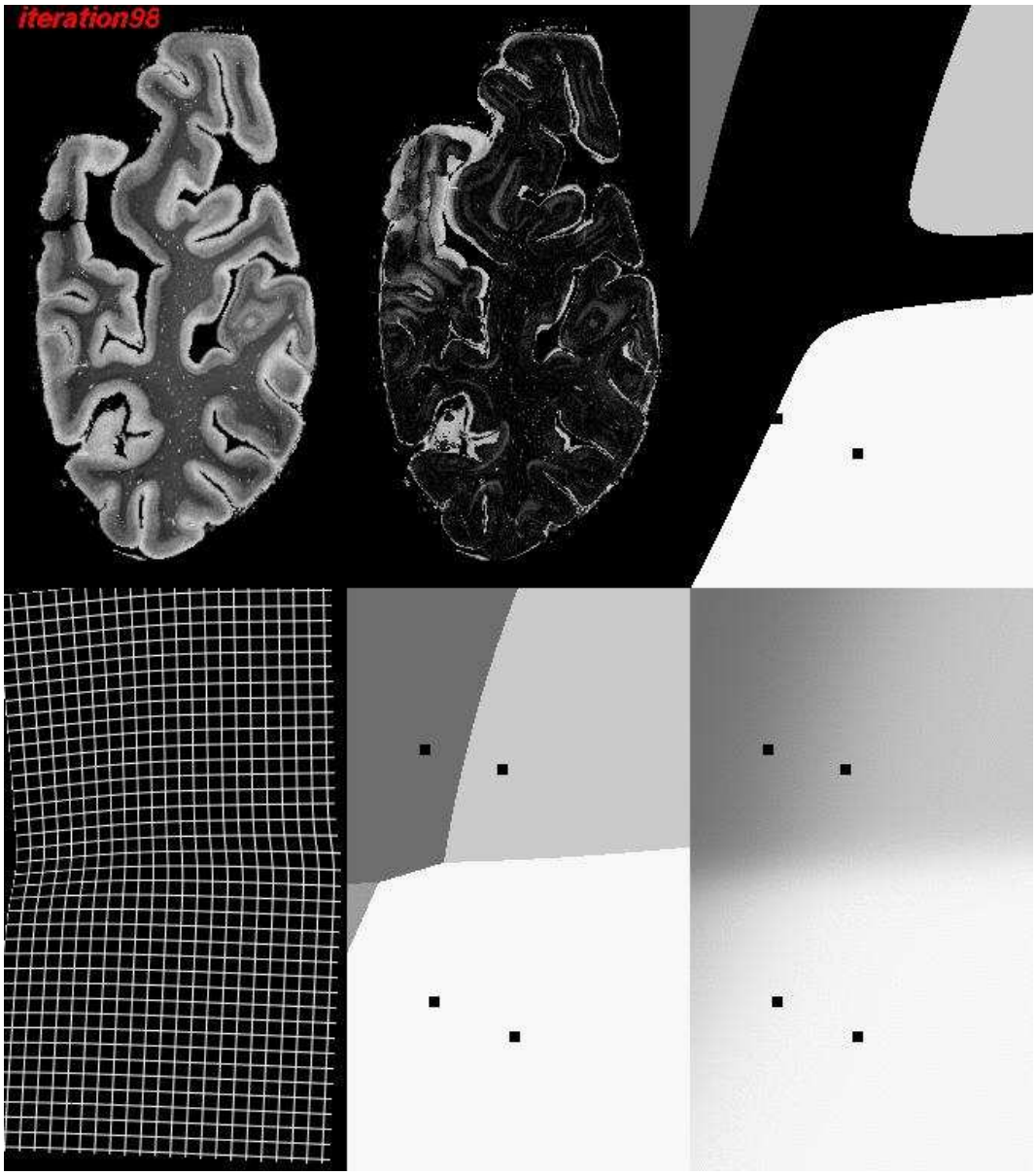

Fig. 7. **Polyrigid registration result with a simple first-order gradient descent.** From left to right and from top to bottom: (1) The deformed image. (2) The image of absolute difference between the deformed image and the fixed image. (3) A representation of the regions of influence: a grey level is attributed to each region, and this color is displayed if and only if the local weight of the region represents more than 90 percents of the total weight. The anchor points are represented here by small squares. (4) A regular grid deformed like the deformed image. (5) An image of the regions of influence, a grey level being displayed if and only if its associated weight is the largest one. (6) An image of the regions of influence displaying at each point the weighted average of the grey levels according to the local weights. Thus, we see that a non-linear deformation has been obtained, as show the curved lines of the initially regular grid. The defect of this registration is that anchor points have not moved from their initial positions.

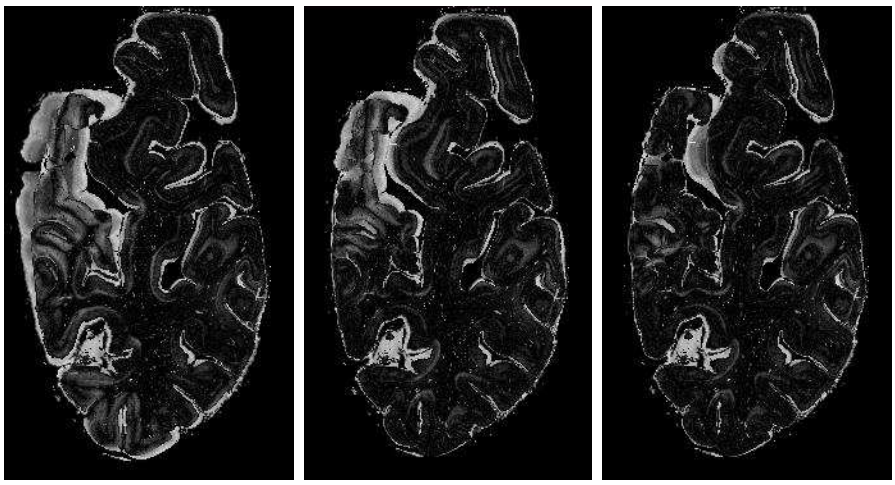

Fig. 8. **Rigid (on the left) versus polyrigid registration with a simple first-order gradient descent and a rescaling for rotations (on the middle) and Levenberg-Marquardt second-order descent(on the right).** This demonstrates that the absence of matter in the lower left-hand corner has thwarted the polyrigid registration algorithm. In the case of the simple first-order gradient descent, most edges have been very finely registered, much better than in the rigid case. Nonetheless, we see that the gyrus lying in the upper left-hand corner still has not been completely correctly registered, because of the incapacity of the algorithm to optimize the region parameters, which have small derivatives in magnitude. On the contrary, on the right, the Levenberg-Marquardt method has allowed the algorithm to register the previously rotated gyrus. The result is much better qualitatively than for the first order descent, and edges are much more finely registered than in the rigid case. However, some amount of unnatural deformations has been added at the vertical frontier between the gyrus and the rest of the slice. This phenomenon is due to the simple forms of the respective rigid regions.

local rigid transformations have been correctly identified. The edges have been very finely registered as compared to rigid registration, as we see in Figure 8. This good result is obtained in spite of a very crude initialization which proves the robustness of the proposed registration algorithm. The only remaining problem is the large deformations occurring at the vertical frontier between the originally rotated gyrus and the other gyri. This is partly due to the simple spherical form chosen for the regions of influence, and partly to the discontinuity that originally made the gyrus rotate. The polyrigid deformations are smooth transformations and therefore they cannot properly model discontinuous deformations.

4.4 Alternating Optimization

The renormalizing process via a second-order approach can be avoided by simply optimizing the parameters alternatively. Moreover, with more than 4

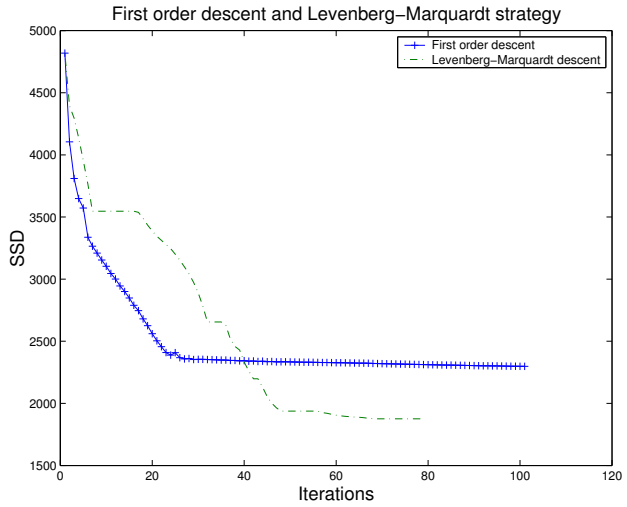

Fig. 9. **SSD criterion evolution for a polyrigid registration with a Levenberg-Marquardt (LM) versus a simple first-order optimization scheme.** This shows that using a second-order descent has greatly enhanced the final results quantitatively, and also qualitatively as is shown is the next figures.

rigid components, the renormalization introduced in the second-order descent is no longer sufficient: the same defects as in the first-order gradient descent appear again.

As a consequence, we introduce here a strategy optimizing alternatively the various parameters. There is no single way of optimizing alternatively the parameters, and it is theoretically difficult to decide which parameters to group, and how many iterations of optimization are to be used for each group at each iteration of the global optimization. Our tests led us to optimize on the one hand the deformation parameters and on the other the region parameters, one iteration at a time for each. We also use here a Levenberg-Marquardt strategy for each group, to speed up the convergence. This yields a stable and efficient optimization algorithm.

Fig. 10 shows the result of the registration. We can clearly see that the registration process has identified and satisfactorily estimated at least three independent rigid behavior. At the same time, the deliberate simplicity of the regions of influence forbids a precise description of the frontiers between the regions. At this point of the registration process, we could resort to a classical non-rigid registration algorithm to make the registration more precise in this sector. But more simply, we can make use of the flexibility of polyrigid transformations by refining the parameters describing the regions of influence, as is shown in the next section of this report.

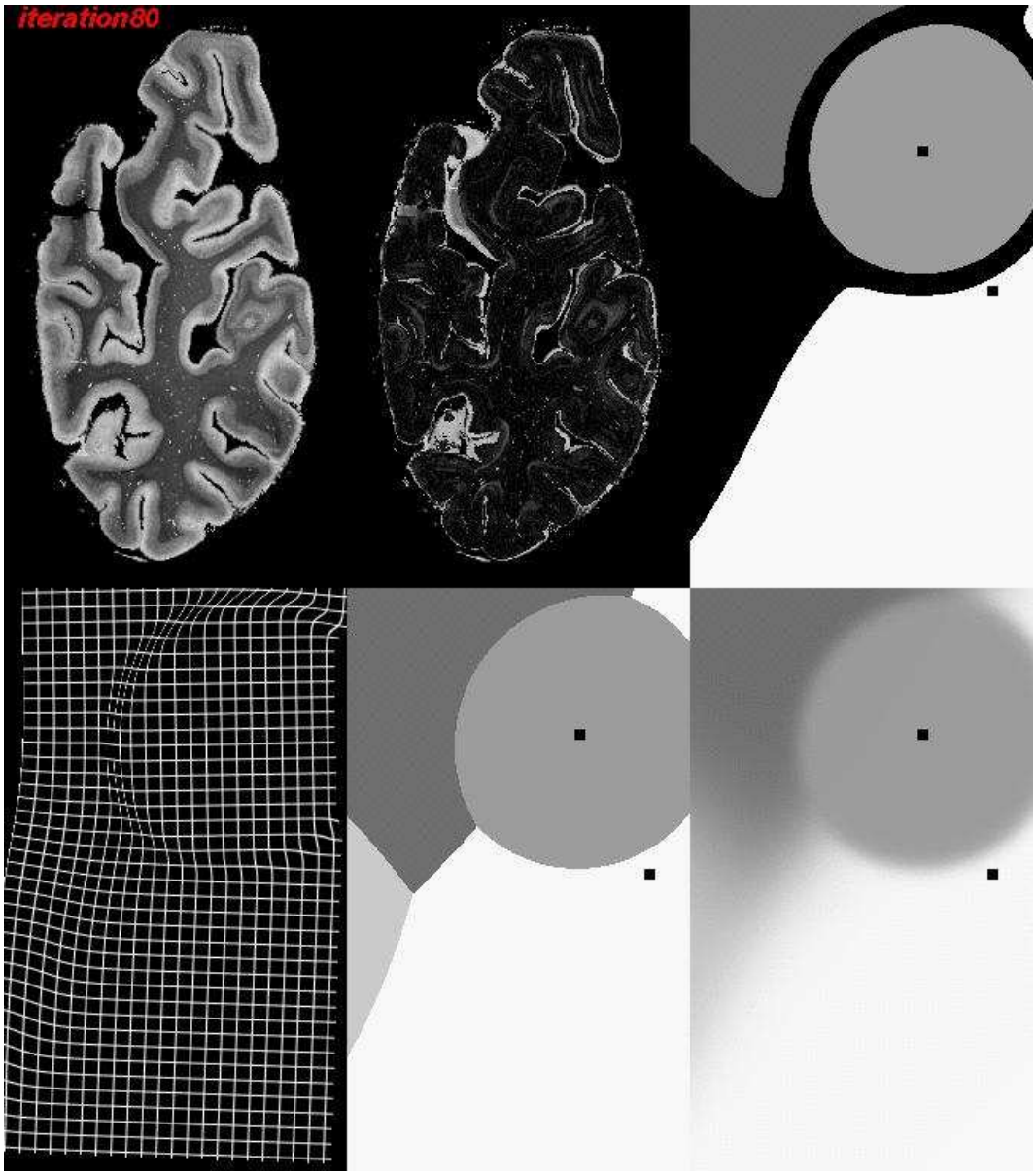

Fig. 10. **Polyrigid alternating LM registration.** The gyrus has been as correctly registered as can be. Due to the Gaussian model, the vertical frontier on the left of it has a circular form, which results in some unnatural deformations. These deformations are marginal but nonetheless non-negligible. However, only 4 rigid components (i.e. 27 scalar parameters) have been necessary here to register very finely most of the slices, without being disturbed by the lack of matter in the lower-left-hand corner of one the registered slice.

5 Preliminary Results with more Complex Regions

5.1 The Shape of the Regions of Influence.

The assumption that each fuzzy region can be accurately described by a simple Gaussian weight can be too strong in certain cases. But generally speaking,

we can be very flexible with the weights because the only limitation is to keep the weights (strictly) positive and smooth with respect to spatial coordinates and parameters. Therefore very complex regions can be used, the simplest way being to use mixtures of simple probability distributions. But other solutions could be used, such as introducing explicitly an pre-defined shape for a region. More precisely, if R is a region, we can define an associated weight with $w(x) = \mathbf{1}_R \star G_\sigma(x)$. $\mathbf{1}_R$ is simply the function returning 1 if $x \in G$ and 0 elsewhere. G_σ is a Gaussian of standard deviation σ, that smoothes $\mathbf{1}_R$ through a convolution. Thus, combinations of pre-defined regions and simply parameterized regions provides quite a rich framework for modeling an application-specific polyrigid transformation.

We present here preliminary results in which we have simply increased the number of anchor points per region. Therefore, regions are modeled via a mixture of Gaussian. This more general form for the weights $w_i(x)$ can be written as follows:

$$w_i(x) = p_i \sum_{j=1}^{n_i} G_{(a_i^j, \sigma_i^j)}(x).$$

In other terms, each component i has its own number n_i of anchor points $(a_i^j)_{j \in 1 \cdots n_i}$, which all have their specific standard deviation σ_i^j.

5.2 Results Obtained with Three Anchor Points

In order to see whether we can obtain better results than in the previous section, we present here the results with three anchor points per region. One could think of refining progressively the number of points, and this is a issue that will be addressed in future work. The present experiment simply consists in making the whole registration proceed with three anchor points, using the most efficient optimization algorithm presented in this report, i.e. the alternating LM strategy.

The experimental setup is identical, except for anchor points, which are initialized on the vertices of equilateral triangles placed on a regular grid.

We obtain here much better results, as show Fig. 11. The frontier that was lacking in precision is substantially refined here, introducing less artificial deformations. However, some amount of unrealistic deformation remains. That was to be expected, since it was because of a rift that the gyrus rotated. To proceed further, it would be necessary to make a distinction between the empty background and matter. A possibility would be to add this knowledge in the weights defining the influence of each region, for instance with a geodesic distance. The weight of a region would then be all the stronger as the current point is close in some geodesic sense. This would result in a different smooth-

ing. The influence of a region close spatially but separated from the current point by a rift of empty background would be seriously lessened. As a consequence, the frontiers between the regions would be more realistic, since they would bear more resemblance with the rifts where they appear.

6 Conclusion and Perspectives

We present in this paper a novel and innovative type of geometrical transformation, the polyrigid and polyaffine transformations. These transformations have several rigid or affine components, which means that a given number of fuzzy regions are defined, on which the global transformation is mostly rigid or affine. The parameters coding the transformation are simple and intuitive, and provide a compact representation for locally rigid or affine movements.

In a rigorous mathematical framework, we show that these transformations are smooth and invertible. We design a new and efficient numerical scheme for the practical implementation in any dimension in the polyrigid case, and devise a complete optimization strategy for its use in the non-rigid registration of medical images (Sec. 4).

Polyrigid transformations are exemplified successfully on the 2D registration of histological slices. Most non-linear artifacts generated during the acquisition process of the slices have been corrected, and it remains only a residual deformation due to the smoothness of polyrigid transformations. For this specific application, further developments would be needed to model the tearing process that has taken place, which is discontinuous by nature.

As shown in Sec. 5, there are many ways of adapting the polyrigid transformations to new applications, by modifying the shape and parameterization of the regions of influence. In order to make the polyrigid transformations more accurate, it should also be possible to define adaptive strategies progressively refine the shape of regions where it is necessary.

We will investigate in future work the application of this new tool to 3D registration. In the human body, many structures present articulations between rigid structures, which suggests the use of transformations incorporating all these rigid movements. A possibility would be to use several components of elongated shape in order to model articulated regions, plus another one modeling the transformation of the background. Such a model has the advantage of accurately describing very complex and non-rigid movements with a limited but adequate number of degrees of freedom.

We have also presented in Sec. 2 the extension of our framework to polyaffine

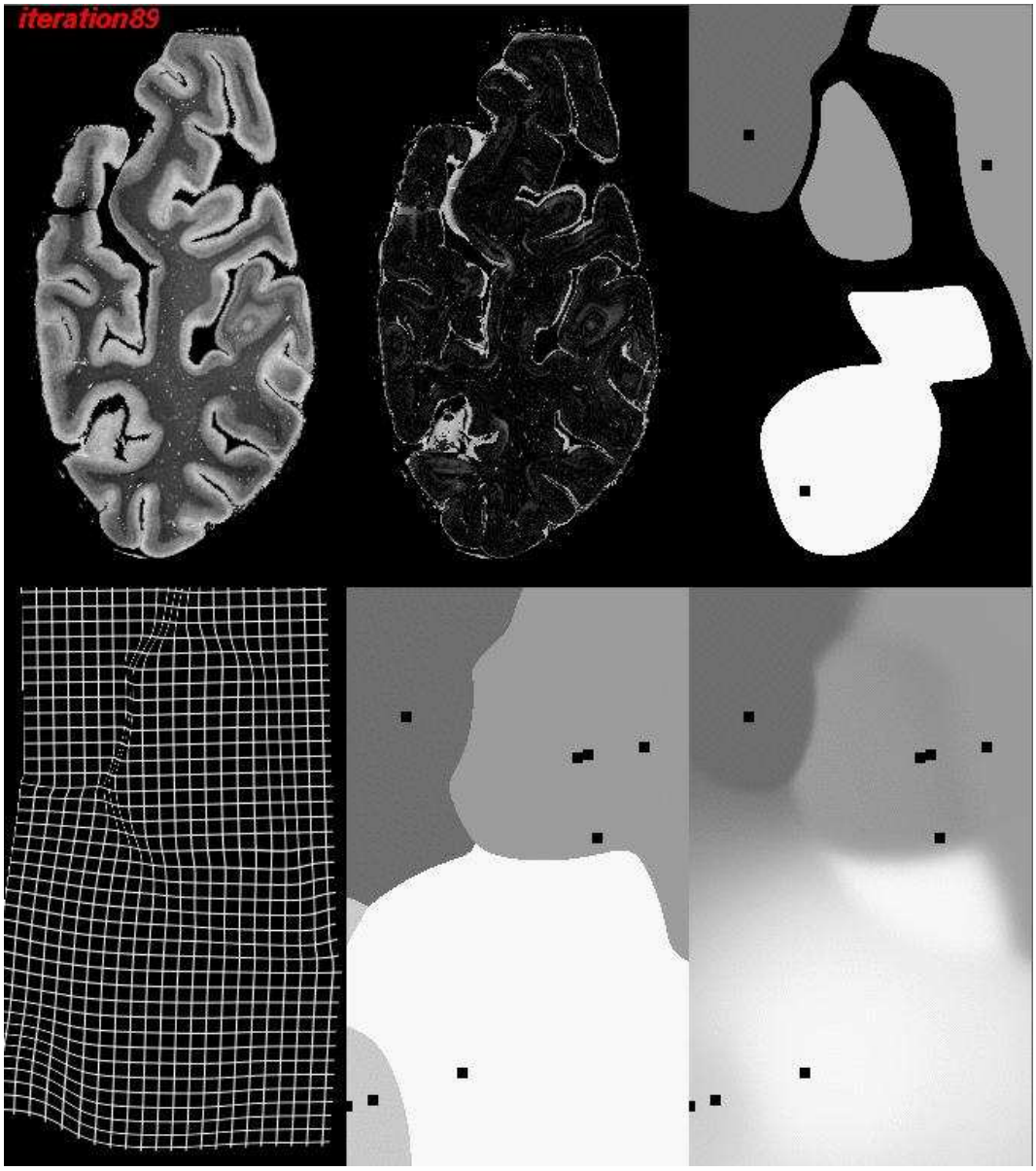

Fig. 11. **Polyrigid alternating LM registration with 3 anchors points per region.** The result is quite satisfactory: thanks to the Gaussian mixture model, a realistic frontier has been automatically inferred which brings the originally rotated gyrus into a precise registration. All edges have been correctly registered. Few artificial deformations are introduced, thanks to the fact that we have only used four different regions having independent rigid motions. As the deformations of the regular grid show, the transformation is still invertible. It should also be noted that this result has been obtained with a fully automatic and crude initialization, and without resorting to a multi-resolution framework. This demonstrates the robustness of the registration algorithm.

transformations. We believe it is possible to use such an extension in the field of shape statistics. More precisely, one could model the variability of the shape around its mean via the statistical analysis of these variations in a certain space of transformations. By choosing as adequately as possible this space of deformations, a model with a limited number of parameters could be derived. Polyaffine transformations are in our opinion a good candidate for doing so, because they can take into account both local rotations, translations or swellings.

In the same vein, another application would be the building of new anatomical atlases, in the case of dataset presenting rigid by part deformations. Using adapted transformations to establish correspondences between the various instances would surely lead to more accurate results. In the case of local swellings or shearings, it would be interesting to compare the performances of these new transformations to those obtained for example with B-Splines, for an equal number of degrees of freedom.

Acknowledgments

We are very grateful to P. Thompson, A. Toga, J. Annese and A. Pitiot for providing us with the histological slices through an associated teams collaboration between Epidaure at INRIA and LONI at UCLA.

A First Derivative of Polyrigid Transformations

Here, we will only focus in 3D on the derivatives of the second scheme, which is the most efficient.

A.1 Differentiation with Respect to Parameters

Let us denote:

$$M_i^{1/N}(x,s) = \frac{1}{N}t_i + (\exp(A_i/N) - Id)\,(x - st_i).$$

This is the modification "proposed" by the i-th component at a given time s and point x for the second scheme. Conversely, let us write the real modification:

$$M^{1/N}(x,s) = \frac{\sum_i w_i(x)(\frac{1}{N}t_i + (\exp(A_i/N) - Id)\,(x - st_i))}{\sum_i w_i(x)}.$$

Then let p_i be a parameter of a rigid transformation T_p, and more specifically a parameter of the i-th component. When we compute the derivative of $T_2^{1/N}(x,s)$ with respect to p_i, we get the following simplification:

$$\begin{aligned}
\frac{\partial T_2^{1/N}(x,s)}{\partial p_i} &= \frac{\partial}{\partial p_i}\frac{\sum_j w_j(x)\left(\frac{1}{N}t_j + (\exp(A_j/N) - Id)(x - st_j)\right)}{\sum_j w_j(x)} \\
&= \frac{\frac{\partial w_i}{\partial p_i}(x)M_i^{1/N}(x,s)}{\sum_j w_j(x)} + \frac{w_i(x)\frac{\partial}{\partial p_i}M_i^{1/N}(x,s)}{\sum_j w_j(x)} \\
&\quad - \left(\frac{\sum_j w_j(x)M_j^{1/N}(x,s)}{\sum_j w_j(x)}\right)\left(\frac{\frac{\partial}{\partial p_i}(\sum_j w_j(x))}{\sum_j w_j(x)}\right) \\
&= \frac{\frac{\partial w_i}{\partial p_i}(x)}{\sum_j w_j(x)}\left(M_i^{1/N}(x,s) - M^{1/N}(x,s)\right) + \frac{w_i(x)}{\sum_i w_i(x)}\frac{\partial}{\partial p_i}M_i^{1/N}(x,s).
\end{aligned}$$

Then, it only remains to see what form take the derivatives of the modifications and of the weights. If we assume that weights have a Gaussian expression as follows:

$$w_i(x) = \frac{p_i}{(2\pi\sigma_i^2)^{n/2}}\exp\left(-\frac{\|x-a_i\|^2}{2\sigma_i^2.}\right). \tag{A.1}$$

Then, we obtain:

$$\begin{cases}
\frac{\partial}{\partial a_i}(w_i(x)) = -\frac{w_i(x)}{\sigma_i^2}(a_i - x)^T. \\
\frac{\partial}{\partial \sigma_i}(w_i(x)) = (-n)\frac{p_i}{(2\pi)^{n/2}\sigma_i^{n+1}}\exp\left(-\frac{\|x-c_i\|^2}{2\sigma_i^2}\right). \\
\qquad\qquad + \frac{p_i}{(2\pi)^{n/2}\sigma_i^n}\frac{\|x-a_i\|^2}{\sigma_i^3}\exp\left(-\frac{\|x-a_i\|^2}{2\sigma_i^2}\right). \\
\qquad\qquad = \left(-\frac{n}{\sigma_i} + \frac{\|x-a_i\|^2}{\sigma_i^3}\right)w_i(x). \\
\frac{\partial}{\partial p_i}(w_i(x)) = \frac{1}{(2\pi\sigma_i^2)^{n/2}}\exp\left(-\frac{\|x-a_i\|^2}{2\sigma_i^2}\right).
\end{cases} \tag{A.2}$$

For derivatives of the modifications, we have:

$$\frac{\partial}{\partial t_i}M_i^{1/N}(x,s) = \frac{1}{N}Id - s\ (\exp(A_i/N) - Id). \tag{A.3}$$

It remains to be seen how one can differentiate $(\exp(A_i/N) - Id)$ with respect to the rotation vector r_i.

A.2 Differentiation with Respect to the Rotation vector

The computation of the derivative of a matrix exponential of a matrix function has no simple form as in the case of scalars. Indeed, when we take $M(p) = \exp(A(p))$, we do not have in general $\frac{\partial}{\partial p}M(p) = \{\frac{\partial}{\partial p}A(p)\}M(p)$. This comes from the non-commutation of $A(p)$ and $\frac{\partial}{\partial p}A(p)$, which is a sufficient condition for differentiating in a simple way the exponential.

Let us denote B_x, B_y, B_z the following matrices:

$$B_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \; B_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \; B_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We have the following result:

$$\forall a \in \{x,\ y,\ z\},\ \frac{\partial}{\partial r_a} \exp\left(A/N\right) = \sum_{n>0} \frac{1}{n! N^n} \sum_{i=1}^{n} A^{i-1}\, B_a\, A^{n-i}.$$

This simply stems from the differentiation of each term of the series defining the exponential.

A.3 Spatial Derivatives

Finally, let us consider the spatial derivative of our scheme, which one must also compute in order to obtain the derivative of the transformation with respect to its parameters. We have:

$$\begin{aligned} \frac{\partial T_p^{1/N}(x,s)}{\partial x} &= \frac{1}{N} \frac{\sum_i (\, M_i^{1/N}(x,s)\, \frac{\partial w_i(x)}{\partial x} + w_i(x) \frac{\partial}{\partial x} M_i^{1/N}(x,s))}{\sum_i w_i(x)} \\ &\qquad - \frac{(\sum_i w_i(x)\, M_i^{1/N}(x,s))(\sum_i \frac{\partial w_i(x)}{\partial x})}{(\sum_i w_i(x))^2} \\ &= \frac{1}{N} \frac{\sum_i (\, M_i^{1/N}(x,s)\, \frac{\partial w_i(x)}{\partial x} + w_i(x)(e^{\frac{A_i}{N}} - Id))}{\sum_i w_i(x)} \\ &\qquad - M^{1/N}(x,s) \frac{(\sum_i \frac{\partial w_i(x)}{\partial x})}{(\sum_i w_i(x))}. \end{aligned} \tag{A.4}$$

And finally, the spatial derivative of the weights is given by:

$$\begin{aligned}\frac{\partial w_i(x)}{\partial x} &= \frac{\partial}{\partial x}\left(\frac{p_i}{(2\pi\sigma_i^2)^{n/2}}\,\exp\left(-\frac{\|x-a_i\|^2}{2\sigma_i^2}\right)\right)\\ &= \frac{p_i}{(2\pi\sigma_i^2)^{n/2}}\left(-\frac{1}{\sigma_i^2}(x-a_i)^T\right)\exp\left(-\frac{\|x-a_i\|^2}{2\sigma_i^2}\right) \qquad \text{(A.5)}\\ &= -\frac{w_i(x)}{\sigma_i^2}(x-a_i)^T.\end{aligned}$$

References

Bardinet, E., Ourselin, S., Malandain, G., Tandé, D., Parain, K., Ayache, N., Yelnik, J., 2002. Three dimensional functional cartography of the human basal ganglia by registration of optical and histological serial sections. In: IEEE International Symposium on Biomedical Imaging. Washington, USA, pp. 329–332.

Bazaraa, M. S., Sherali, H. D., Shetty, C. M., 1993. Non linear programming, theory and algorithms, 2nd edition. John Wiley & Sons, Inc.

Bookstein, F. L., 1999. Linear methods for nonlinear maps: Procrustes fits, thin-plate splines, and the biometric analysis of shape variability. In: Toga, A. (Ed.), Brain Warping. Academic Press, pp. 157–181.

Cachier, P., 2002. Recalage non-rigide d'images médicales volumiques, contributions aux approches iconiques et géométriques. Ph.D. thesis, École Centrale Paris.

Cachier, P., Bardinet, E., Dormont, D., Pennec, X., Ayache, N., Feb.-march 2003. Iconic feature based nonrigid registration: The PASHA algorithm. CVIU — Special Issue on Nonrigid Registration 89 (2-3), 272–298.

Camion, V., Younes, L., 2001. Geodesic interpolating splines. In: Figuiredo, M., Zerubia, J., Jain, A. (Eds.), Proc. of Energy Minimization Methods in Comp. Vis. and Pat. Rec. (EMMCVPR'01). LNCS 2134. pp. 513–527.

Chefd'hotel, C., Hermosillo, G., Faugeras, O., Jul. 2002. Flows of diffeomorphisms for multimodal image registration. In: Proc. of IEEE Int. Symp. on Biomedical Ima. Washington D.C.

Ferrant, M., Warfield, S., Guttmann, C., Mulkern, R., Jolesz, F., Kikinis, R., 1999. 3D image matching using a finite element based elastic deformation model. In: Proc. of MICCAI'99. LNCS 1679. pp. 202–209.

Hermosillo, G., Chefd'Hotel, C., Faugeras, O., December 2002. Variational methods for multimodal image matching. IJCV 50 (3), 329–343.

Little, J., Hill, D., Hawkes, D., May 1996. Deformations incorpotationg rigid structures. Computer Vision and Image Understanding 66 (2), 223–232.

Maintz, J., Viergever, M., 1998. A survey of medical registration. Medical image analysis 2 (1), 1–36.

Ourselin, S., Bardinet, E., Dormont, D., Malandain, G., Roche, A., Ayache, N., Tande, D., Parain, K., Yelnik, J., October 2001a. Fusion of histological

sections and MR images: towards the construction of an atlas of the human basal ganglia. In: Niessen, W., Viergever, M. (Eds.), 4th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'01). Vol. 2208 of LNCS. Utrecht, The Netherlands, pp. 743–751.

Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N., Jan. 2001b. Reconstructing a 3D structure from serial histological sections. Image and Vision Computing 19 (1-2), 25–31.

Pitiot, A., Malandain, G., Bardinet, E., Thompson, P., 2003. Piecewise Affine Registration of Biological Images. In: Second International Workshop on Biomedical Image Registration WBIR'03.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., Hawkes, D. J., 1999. Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Medecal Imaging 18 (8), 712–721.

Sheppard, D., 1968. A two-dimensionnal interpolation function for irregularly spaced data. In: 23rd National Conference of the ACM. ACM Press, pp. 517–524.

Tenenbaum, M., Pollard, H., 1985. Ordinary Differential Equations. Dover.

Thirion, J.-P., 1998. Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis 2 (3), 243–260.

Wüstner, M., 2003. A connected lie group equals the square of the exponential image. Journal of Lie Theory 13, 307–309.