A Forward Model to Build Unbiased Atlases from Curves and Surfaces

Abstract. Building an atlas from a set of anatomical data relies on (1) the construction of a "mean anatomy" (called template or prototype) and (2) the estimation of the variations of this template within the studied population. To avoid biases introduced by separate processing, we *jointly* estimate the template as well as its deformation on the basis of a consistent statistical model. We use here a forward model that considers data as noisy deformations of an unknown template. This differs from backward schemes which estimate a template by pulling back data into a common reference frame. If we need here to introduce specific numerical schemes to build atlases from curves and surfaces, the comparison of any new available data with the learnt model should be simpler in this case than in the backward setting. We show the feasibility of the approach by building atlases from sets of 34 sulcal lines and 50 deep brain structures.

1 Forward vs. Backward Models for Template Estimation

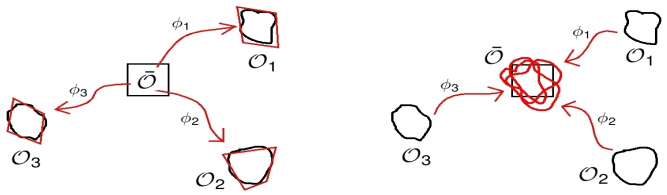
In the medical imaging field, atlases are useful to drive the personalization of generic models of the anatomy, to analyze the variability of an organ, to characterize and measure anatomical differences between groups, etc. Many frameworks have been proposed to build atlases from large database of medical images [1,2,3,4], much fewer were proposed for anatomical curves or surfaces [5,6]. In any case, the underlying idea remains the same: one estimates a "mean anatomy" (called template) and one learns how this mean model deforms within a given population. The most widely used method in medical imaging is based on a *backward* model that deforms every observations back to a common reference frame (See Fig.1). However, we prefer here to base our statistical estimation on a *forward* model, as pioneered in [7,8], which considers the observations (T_i) as noisy deformations (ϕ_i) of an unknown template ($\bar{T}$). Formally, the *forward* model can be written as:

$$T_i = \phi_i.\bar{T} + \varepsilon_i \tag{1.1}$$

whereas the *backward* model is:

$$\phi_i.T_i = \bar{T} + \varepsilon_i \quad \Longleftrightarrow \quad T_i = \phi_i^{-1}\bar{T} + \phi_i^{-1}\varepsilon_i \tag{1.2}$$

The backward model considers either (Eq.1.2-left) that the template $\bar{T}$ is noisy and the observations T_i free of noise or (Eq.1.2-right) that the noise added the observations ($\phi_i^{-1}\varepsilon_i$) **depends** on the observations via an unknown deformation. By contrast, the forward model (Eq.1.1) considers that the template is not blur, as an "ideal" object, and an independent and **identically distributed**



forward scheme backward scheme

Fig. 1. Two statistical frameworks: in the forward scheme (left), the physical observations ($\mathcal{O}_i$) are seen as noisy deformation (ϕ_i) of unknown template ($\bar{\mathcal{O}}$). In the backward scheme, the template is an average of deformed observations. The two schemes are not equivalent: in the forward scheme the noise is removed from the observations whereas the noise is pulled back in the common reference frame with the backward scheme.

noise ε_i is added to every observations. This follows more accurately the real physical acquisition process, whereas the backward model relies on less realistic assumptions.

The observations T_i are given as discrete sampled objects. The template $\bar{T}$ models an average "ideal" biological material and it is therefore supposed to be continuous. Since in the backward model sampled observations deform to a continuous template, an extrinsic **interpolation** scheme is required. By contrast, in the forward setting, the deformed template needs only to be **sampled** to be compared to the observations. This does not only reproduce more accurately the real physical acquisition process, but also depends on less arbitrary assumptions.

Assume now that we can define probabilities on objects T (images, curves, surfaces, etc.) and on deformations ϕ. The statistical estimation of an atlas would require at least to compute $p(\bar{T}|T_i)$, the probability of having the template given a learning database of T_i. Once the atlas is built, one would like then to know how a new observation T_k is compared to the learnt variability model: one needs to compute the likelihood of this observation given the template $p(T_k|\bar{T})$. Because ϕ_i acts differently in Eq.1.1 and in Eq.1.2, the computational cost of these two steps varies significantly. In the backward scheme, computing $p(\bar{T}|T_i)$ should be simpler than computing $p(T_i|\bar{T})$ which depends on the Jacobian of ϕ_i. It is exactly the reverse for the forward scheme. Since it is better to spend more time to build the atlas (which is done once for all) and to keep simple the test of any new available data, the forward model seems better suited even from a computational point of view.

Finally, the forward model is also better understood from a theoretical point of view. For instance, the convergence of the Maximum A Posteriori (MAP) template estimation, when the number of available observations is growing, is proved for images and small deformations [7,9]. Such proofs for the backward model seem currently out of reach.

For all these reasons, we base here our statistical estimations on the forward model. We show in this paper how the atlas building step, which is the most critical step in this paradigm, is possible in case of curves and surfaces. Com-

pared to images, dealing with shapes requires specific numerical scheme. We take advantage here of a sparse deconvolution scheme we introduced recently [10].

The paper is organized as follows. A general non parametric framework for shape statistics is introduced in Section 2. In Section 3 we detail the optimization procedure to estimate jointly a template and its deformations to the shapes. A sparse deconvolution method is presented to effectively compute the gradient descent. In Section 4, we show templates computed from large sets of sulcal lines and sets of meshes of sub-cortical structures.

2 Non-parametric Representation of Shapes as Currents

As emphasized in [11,12,5], a current is a convenient way to model geometrical shapes such as curves and surfaces. The idea is to characterize shapes via vector fields, which are used to probe them. A surface S is characterized by the flux of any vector field ω through it ($S(\omega) = \int_S \omega(x).(u \times v)(x)d\sigma(x)$ where $u \times v$ is the normal of the surface, (u, v) an orthonormal basis of its tangent plane), a curve L by the path integral of ω along it ($L(\omega) = \int_L \omega(x).\tau(x)dx$ where τ is the tangent of the curve). More generally, we define a current T as a linear continuous mapping from a set of *test* vector fields W to $\mathbb{R}$. This framework enables to define addition, deformation, Gaussian noise on shapes, to measure a distance between shapes, without assuming any point correspondences between objects. We recall here some properties of currents to give a rigorous sense of Eq.1.1 in case of shapes. We refer the reader to [11,13,10] for more details.

As mappings from W to $\mathbb{R}$, the currents build a **vector space**, denoted W^*: $(T_1 + T_2)(\omega) = T_1(\omega) + T_2(\omega)$ and $(\lambda.T)(\omega) = \lambda.T(\omega)$. For surfaces, this means that the flux through two surfaces is the sum of the two flux through each surface: the addition is equivalent to the union of surfaces. Scaling a surface means scaling the power of the flux through the surface.

Suppose now that we can provide the *test* space W with a norm that measures the regularity of the vector fields. We can define then a **norm of a current** T as the maximum flux through T of any regular vector fields (i.e. $\|\omega\|_W \leq 1$): $\|T\|^2_{W^*} = \mathrm{Sup}_{\|\omega\|_W \leq 1} |T(\omega)|$. The distance between two surfaces $\|S - S'\|_{W^*}$ is therefore obtained for the regular vector field that best separate the two surfaces, in the sense that the difference of the flux through the two surfaces is the largest possible one. *This distance between shapes does not depend on how shapes are parametrized and does not assume point correspondences between shapes.*

For computational purposes, we suppose, from now onwards, that W is a reproducible kernel Hilbert space (r.k.h.s) with kernel K [14]. In this setting, the set of all delta Dirac currents δ^α_x is **a basis** of the vector space of currents. A Dirac current δ^α_x is defined by: $\delta^\alpha_x(\omega) = \langle \omega(x), \alpha \rangle_{\mathbb{R}^3}$ for any $\omega \in W$. It may be seen as a tangent (resp. normal) α entirely concentrated at point x. Although a curve (resp. a surface) has an infinite number of tangents (resp. normals), a polygonal lines (resp. a mesh) may be approximated in the space of currents by a finite sum $\sum_k \delta^{\alpha_k}_{x_k}$ where x_k is the center of the segment (resp. center of the mesh cell) and α_k the tangent of the line (resp. the normal of the surface) at x_k.

Moreover, thanks to the Riesz-Frechet theorem, we can show that for any current T, the vector field that realizes the supremum $\mathrm{Sup}_{\|\omega\|_W \leq 1} |T(\omega)|$ exists and is unique. We denote this vector field $\mathcal{L}_W^{-1}(T)$ and call it the **dual representation** of T. We can show that $\mathcal{L}_W$ is isometric [10]: it provides W^* with an **inner product**: $\|T\|^2_{W^*} = \langle T, T \rangle_{W^*} = \left\langle \mathcal{L}_W^{-1}(T), \mathcal{L}_W^{-1}(T) \right\rangle_W$. On basis vectors we have: $\mathcal{L}_W^{-1}(\delta_x^\alpha)(y) = K(y,x)\alpha$ (K is the Green kernel of $\mathcal{L}_W^{-1}$) and $\left\langle \delta_x^\alpha, \delta_y^\beta \right\rangle_{W^*} = \alpha^t K(x,y)\beta$. This gives by linearity *explicit* and *easily tractable* formula to compute the inner product (and then the distance) between two shapes $T = \sum_i \delta_{x_i}^{\alpha_i}$ and $T' = \sum_j \delta_{y_j}^{\beta_j}$: $\langle T, T' \rangle_{W^*} = \sum_i \sum_j \alpha_i^t K(x_i, y_j)\beta_j$. The mapping $\mathcal{L}_W^{-1}$ is a simple convolution operator, as illustrated in Fig.2-a. On the contrary, finding the current T whose associated vector field γ is given ($T = \mathcal{L}_W(\gamma)$) is an ill-posed deconvolution problem that requires specific numerical scheme as the one we will present in Section 3.3.

The kernel K enables also to define an associated **Gaussian noise** on currents ε. The covariance between two coordinates of ε in the basis (δ_x^α) is given by: $\mathrm{Cov}(\langle \varepsilon, \delta_x^\alpha \rangle, \langle \varepsilon, \delta_y^\beta \rangle) = \alpha^t K(x,y)\beta$. For our applications we will choose a Gaussian kernel: $K(x,y) = \exp(-\|x-y\|^2 / \lambda_W^2)$.

To make sense of the model equation Eq.1.1, one must still specify how this modelling based on currents may be coupled with a deformation framework. If ϕ is diffeomorphism and S a surface, the flux of ω through the deformed shape $\phi(S)$, denoted for general currents $(\phi_\sharp \star S)(\omega)$, is equals to the flux through S of the pulled-back vector field $\phi^\sharp \star \omega$ which is given by the change of variables formula within the flux integral. This enables to define a general **push-forward action of a diffeomorphism** on any currents. This action replaces for curves and surfaces the usual action on images: $(\phi \star I)(x) = I(\phi^{-1}(x))$. This is here slightly more complex since we do not transport points but tangents or normals (differential 1 and 2-forms to be even more precise). On the basis elements, the push-forward action gives: $\phi_\sharp \star \delta_x^\alpha = \delta_{\phi(x)}^{d_x\phi(\alpha)}$ if α is a tangent of a curve and $\phi_\sharp \star \delta_x^{u \times v} = \delta_{\phi(x)}^{d_x\phi(u) \times d_x\phi(v)}$ if $u \times v$ is the normal of a surface. One notices that $d_x\phi(u) \times d_x\phi(v) = |d_x\phi| \, d_x\phi^{-t}(u \times v)$.

In the following, we restrict the deformations ϕ to belong to the group of diffeomorphism set up in [15,16]: the diffeomorphisms are obtained by integration of a time-varying vector field v_t: $\partial_t \phi_t = v_t \circ \phi_t$. The geodesic flows $(\phi_t^{v_t})_{t \in [0,1]}$ are completely determined by the initial vector speed v_0 which belongs to a r.k.h.s. V. The final diffeomorphism is denoted ϕ^{v_0}. How to find such a diffeomorphism that best matches two sets of currents is explained in detail in [11,13].

3 Joint Estimation of Template and Deformations

3.1 A Heuristic Maximum A Posteriori in Infinite Dimension

From a Bayesian point of view, in Eq.1.1 ($T_i = \phi_{i,\sharp} \star \bar{T} + \varepsilon_i$), T_i are the observations, $\bar{T}$ is unknown, ϕ_i are hidden variables and ε_i independent and identically

distributed Gaussian noise with known variance. Suppose now that we can define Gaussian probability density functions (pdf) on the space of Currents W^* and on the space of initial vector fields V: $p_\varepsilon(\varepsilon) = C_\varepsilon \exp(-\|\varepsilon\|^2_{W^*}/\sigma^2_W)$ and $p_\phi(v) = C_\phi \exp(-\|v\|^2_V/\sigma^2_V)$. In that case, a Maximum A Posteriori (MAP) estimation for independent observations maximizes $\max_{\bar{T}} \prod_i^N p(T_i|\bar{T})$. Formally,

$$p(T_i|\bar{T}) = \int p_\varepsilon(Ti|\bar{T}, v_0^i) p_\phi(v_0^i) dv_0 = \int p_\varepsilon(T_i - \phi_i^{v_0^i}\bar{T}) p_\phi(v_0^i) dv_0$$

Since the term within the integral depends on v_0 by a geodesic shooting of diffeomorphisms, there are no close forms for this likelihood. A usual approximation consists in replacing the term within the integral by its maximum. This leads to: $p(T_i|\bar{T}) \sim \max_{v_0^i} p_\varepsilon(T_i - \phi_i^{v_0^i}\bar{T}) p_\phi(v_0^i)$ which finally gives:

$$(\bar{T}, \phi^{v_0^i}) = \text{argmin}_{\bar{T}, v_0^i} \left\{ \sum_{i=1}^N \frac{1}{\sigma_V^2} \left\|v_0^i\right\|^2_V + \frac{1}{\sigma_W^2} \left\|T_i - \phi^{v_0^i}.\bar{T}\right\|^2_{W^*} \right\} \tag{3.1}$$

so called, Fast Approximation with Mode (FAM). If we already defined Gaussian variables in the space of currents (Section 2), such variables have no pdf because W^* and V are of infinite dimension. A more rigorous MAP derivation could be done considering finite dimensional parametrization of the v_0's and of the T_i's and Markov Chain Monte Carlo (MCMC) approaches for sampling the posterior could be also possible along the lines of [9] but still challenging and out of the scope of this paper. Here, we minimize directly Eq.3.1.

3.2 A alternated Minimization Procedure

We solve Eq.3.1 by minimizing it alternatively with respect to the template and to the deformations. When the template $\bar{T}$ is fixed, each term of 3.1 can be minimized separately. For a given observation T_i, minimizing $\frac{1}{\sigma_V^2}\left\|v_0^i\right\|^2_V + \frac{1}{\sigma_W^2}\left\|T_i - \phi_i^{v_0^i}.\bar{T}\right\|^2_{W^*}$ with respect to v_0^i is exactly a registration problem, as stated and solved in [11,13]. This step of the minimization consists therefore of N registrations of the template $\bar{T}$ to each observation T_i.

When the deformations ϕ_i are fixed for every $i = 1 \dots N$, minimizing 3.1 with respect to the template $\bar{T}$ leads to the minimization of:

$$J(\bar{T}) = \frac{1}{2} \sum_{i=1}^N \left\|\phi_{\sharp,i} \star \bar{T} - T_i\right\|^2_{W^*} \tag{3.2}$$

If all $\phi_i = \text{Id}$ (i.e. no deformation), the minimum is reached at the empirical mean: $\bar{T} = \frac{1}{N}\sum_i^N T_i$. For arbitrary deformations, there is no close form and we use a gradient descent scheme. The gradient of Eq.3.2 is precisely:

$$\nabla_{\bar{T}} J = \sum_{i=1}^N \phi^*_{\sharp,i} \star (\phi_{\sharp,i} \star \bar{T} - T_i) \tag{3.3}$$

where $\phi_\sharp^*$ is the adjoint action of $\phi_\sharp$: $\left\langle \phi_\sharp^* \star T, T' \right\rangle_{W^*} = \langle T, \phi_\sharp \star T' \rangle_{W^*}$ for any currents T and T'. This would be a matrix transpose if the action were linear. In this non-linear setting, standard computations give $\phi_\sharp^* \star T = \mathcal{L}_W(\phi^\sharp \star \mathcal{L}_W^{-1}(T))$. With the backward scheme, Eq.3.2 would be: $J(\bar{T}) = \frac{1}{2}\sum_{i=1}^N \left\|\bar{T} - \phi_{\sharp,i} \star T_i\right\|_{W^*}^2$, whose minimum has the close form $\bar{T} = \frac{i=1}{N}\sum_1^N \phi_{\sharp,i} \star T_i$. We see here why the atlas building step is computationally more difficult in the forward setting. However, computing the likelihood of any new observations $p(T_i|\bar{T})$ will be much simpler and faster in this setting.

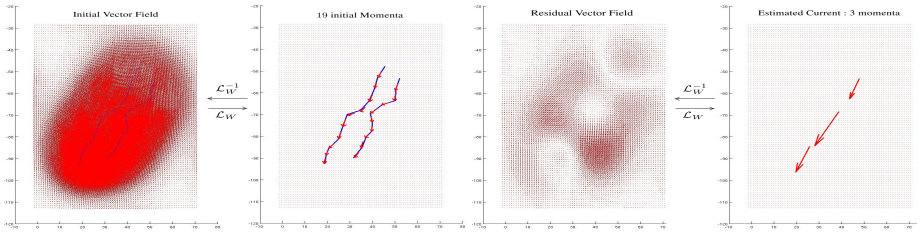
The input shapes T_i are sampled objects which are approximated as finite set of Dirac currents. As it will appear from this minimization procedure, the template will also always remain a finite set of Dirac currents at every iteration. Therefore, the current $\phi_{\sharp,i} \star \bar{T} - T_i$ is of the form $\sum_k \delta_{y_k^i}^{\beta_k^i}$ which finally leads to $\phi_i^\sharp \star \mathcal{L}_W^{-1}(\phi_{\sharp,i} \star \bar{T} - T_i) = \sum_k d_x\phi_i^t K(\phi_i(x), y_k^i)\beta_k^i$ in case of curves. Finally, the vector field associated to the gradient of the energy 3.2 in case of curves can be computed at any point x of the space:

$$\mathcal{L}_W^{-1}(\nabla_{\bar{T}} J)(x) = \sum_{i=1}^N d_x\phi_i^t \left(\sum_k K(\phi_i(x), y_k^i)\beta_k^i \right) \tag{3.4}$$

For surface, $d_x\phi_i^t$ must be replaced by $|d_x\phi|\, d_x\phi^{-1}$. At this stage, we see that the gradient descent scheme could not be performed without an efficient numerical algorithm to estimate the current ∇J whose associated vector field is given by Eq.3.4. The sparse deconvolution method of Section 3.3 precisely provides a finite set of Dirac current which approximates ∇J at any arbitrary precision, so that the template remains a finite set of Dirac currents at each iteration.

We initialize the algorithm by setting $\phi_i = \mathrm{Id}$, $\bar{T} = \frac{1}{N}\sum_{i=1}^N T_i$ and by computing $\gamma_{\bar{T}} = \mathcal{L}_W^{-1}(\bar{T})$ via a Gaussian convolution. The current $\bar{T}$ is encoded as a list of (position, vectors) that approximate small segments or small triangles. The dense vector field $\gamma_{\bar{T}}$ is discretized at a fixed grid's points: $\Lambda = \{x_p\}$ and therefore encoded as an image of vectors. The following variable grad is also an image of vectors. We then iterate the following loop:

- For $i = 1 \dots N$, $\phi_i \leftarrow$ registration of $\bar{T}$ to T_i.
- Until convergence do
  - grad $= 0$
  - For $i = 1 \dots N$ do
    - transport tangents (normals) of $\bar{T}$ with ϕ_i (gives $\phi_{\sharp,i} \star \bar{T}$).
    - store the (y_k^i, β_k^i) such that $\phi_{\sharp,i} \star \bar{T} - T_i = \sum_k \delta_{y_k^i}^{\beta_k^i}$
    - Deform Λ with ϕ_i and for each p compute:
      - $d_{x_p}\phi_i$ by a finite difference scheme
      - $G = \sum_k K(\phi_i(x_p), y_k^i)\beta_k^i$ (convolution)
      - $\mathrm{grad}(p) \leftarrow \mathrm{grad}(p) + 2d_{x_p}\phi_i^t G$ (or $+2\left|d_{x_p}\phi_i\right|(d_{x_p}\phi_i)^{-1}G$)
  - $\gamma_{\bar{T}} \leftarrow \gamma_{\bar{T}} - \tau\mathrm{grad}$
  - Deconvolution of $\gamma_{\bar{T}}$ to give the new $\bar{T}$ (See section 3.3).



a- Mean of two curves b-Approximation with 3 Dirac currents

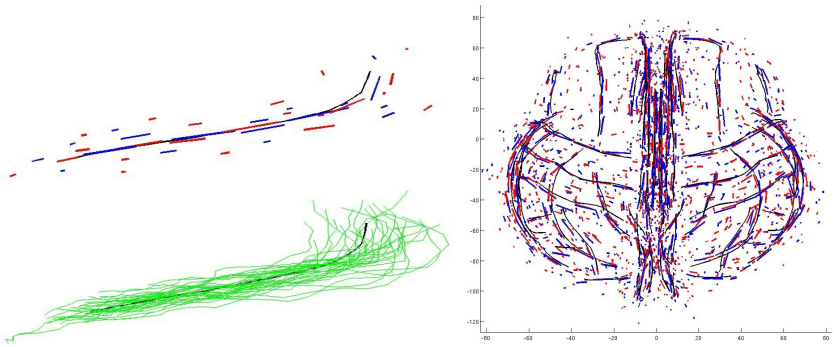
Fig. 2. Toy example in 2D illustrating the sparse deconvolution scheme. **a**-right: two initial curves (L_1, L_2 in blue), their mean ($(L_1 + L_2)/2$) is represented by a set of 19 Dirac currents in red. **a**-left: a Gaussian convolution ($\mathcal{L}_W^{-1}$) of the 19 Dirac gives the dual representation of the mean as a dense vector field in W. **b**- The deconvolution method estimates iteratively a set of Dirac currents (**b**-right after 3 steps) whose convolution retrieves the initial vector field with increasing precision. The difference between the true vector field and the estimated one is shown in **b**-left.

3.3 Sparse Deconvolution by a Matching Pursuit Algorithm

A current may be represented by a very large number of Dirac currents as, for instance, the mean current $\bar{T} = \sum_i^N T_i = \sum_k^n \delta_{x_k}^{\alpha_k}$ where n grows linearly with the number of observations. This makes critical the registration of $\bar{T}$ right after the initialization step. But this representation is often far from being optimal: it may be highly redundant at the scale λ_W. To integrate this redundancy, we compute the vector field $\mathcal{L}_W^{-1}(\bar{T})$ (Fig.2-a) by convolution. Then, an adequate deconvolution scheme could be applied to estimate an adapted basis on which the true solution T is decomposed in $T = \sum_{k'} \delta_{x_{k'}}^{\alpha_{k'}}$ with fast decreasing terms. The first terms of this series will give therefore an approximation of $\bar{T}$ with an increasing precision (Fig.2-b).

In Eq.3.4, we make computations on vector fields and we do not know how to retrieve the current T that comes from the resulting vector field. The solution may require theoretically an infinite number of Dirac currents: $T = \sum_{i=0}^{\infty} \delta_{x_i}^{\alpha_i}$. Here again, we need a deconvolution scheme that estimates iteratively the most important terms of this series. The sum of these first terms will provide an approximation of T at any desired accuracy.

The deconvolution scheme we introduced in [10], adapts the orthogonal matching pursuit, as introduced in [17] to decompose images in adapted wavelet bases, to our framework based on currents. Given a vector field $\gamma \in W$, the first step consists in finding the point x_1 that maximizes the projection of $\mathcal{L}_W(\gamma)$ on δ_{x_1}: $\langle \mathcal{L}_W(\gamma), \delta_x^{\varepsilon_k} \rangle_{W^*} = \langle \gamma, K(.,x)\varepsilon_k \rangle_W = \gamma(x)_k$ (x_k is an orthonormal basis of $\mathbb{R}^3$). x_1 reaches therefore the maximum of γ. A linear set of equation determines α_1 such that $\delta_{x_1}^{\alpha_1}$ is the orthogonal projection of $\mathcal{L}_W(T)$ into $\mathrm{Span}(\delta_{x_1})$. We substract then $K(.,x_1)\alpha_1$ from γ and we iterate on this residue. This builds iteratively a series that has been proved in [10] to converge toward the true solution.



a- Sylvian Fissure (Left Hemisphere) b- All 70 sulci

Fig. 3. Estimated template from 34 subjects. Left: initial 34 Sylvian Fissure of the left hemisphere (green), the empirical mean (red) and the estimated template (blue). In black, the mean lines computed from B-spline parametrization of curves [18]. Right: Same curves for all 70 sulci. Although results look similar, only the template in blue is not biased in the sense of the model Eq.1.1.

4 Experimental Results

4.1 Building Atlases from Curves

We use a database of cortical sulcal landmarks (72 per brain) delineated in a large number of subjects scanned with 3D MRI (age: 51.8 $\pm$ 6.2 years). From 34 subjects in the database, we build the template according to our estimation method. We set the parameters $\lambda_W = 12$mm, $\lambda_V = 25$mm and $\gamma = \sigma_W^2/\sigma_V^2 = 0.01$. The diameter of the brain is typically 120mm. Figure 3 shows the estimated template after 2 iterations. During the gradient descent, the bias (in the sense of our forward model Eq.1.1) is removed from the empirical mean (in red), leading to an unbiased template (in blue). The sparse deconvolution scheme enables also to give a light representation of the template: whereas the database contains 32643 segments (on average 960 segments per subject), the estimated template is represented by only 1361 Dirac currents with an approximation error below 5%. This would be of great interest in the future, for example to register this template toward any new available data.

4.2 Building Atlases from Surfaces

We use a database of 10 segmented sub-cortical structures of the brain (Caudate, Putamen, Globus Pallidus, Amygdala and Hippocampus) in large number of autistics and healthy children scanned with 3D MRI (age 2.7 $\pm$ 0.3) [19]. From 25 autistics and 7 controls, we build the template according to our estimation method for surfaces. We use the parameters $\lambda_W = 5$mm, $\lambda_V = 20$mm, $\gamma = \sigma_W^2/\sigma_V^2 = 0.001$. The set of 10 sub-cortical structures have typically a diameter of 60mm. Figures 4-a,b show the estimated template after 2 iterations. Each Dirac

a- Autistics b- Controls c- Difference on hippocampus

Fig. 4. Estimated templates from 25 autistics (a) and 7 controls (b) for 10 sub-cortical structures (one per color). In (c), the blue arrows approximate the difference between autistics and controls' template, superimposed with the hippocampus of a control.

current δ_x^α is represented by an equilateral triangle whose center of mass is x and non-normalized normal α. Each of the 10 meshes has 2880 cells, so that the total number of normals is 720e3. Thanks to the deconvolution scheme, the estimated template is represented by only 1344 normals with an approximation error below 5%. In Fig.4-c, we show the difference between the template of autistics and controls approximated by the proposed matching pursuit algorithm. Note that the anatomical differences between both classes are not only captured by their respective templates but also by the main modes of the deformations ϕ_i.

5 Conclusion and Perspectives

In this paper, we propose a statistical model which estimates jointly a template, the deformations of the template to the observations and the noise on these observations. This differs from methods that use extrinsic template, thus introducing a bias in the analysis of the variability. Our forward model considers observations as noisy deformations of an unknown template. Compared to the more commonly used backward setting, this model requires a specific deconvolution scheme to deal with the non-linear action of diffeomorphisms on curves and surfaces. However, once the model is built, we can decompose any new data into a deformation of the template and a residual noise and then measure how likely this decomposition may be with respect to the model. This offers a way to classify data according to pathologies, to detect outliers in a database, to highlight where a given observation differs from a model, etc. Such a statistical inference would be more difficult within the backward framework since the likelihood of any new observations would involve the Jacobian of the deformations that we take here directly into account while building the atlas.

Numerical experiments show the feasibility and relevance of our approach. In particular, two templates were estimated from deep brain structures of autistics children and controls. The results suggest anatomical differences between both classes. However, such results should be strengthen by rigorous statistical tests.

Our future work will also evaluate the model with respect to its prediction and classification capability.

References

1. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23** (2004) 151–160
2. Avants, B., Gee, J.: Geodesic estimation for large deformation anatomical shape averaging and interpolation. NeuroImage **23** (2004) 139–150
3. Marsland, S., Twining, C.: Constructing diffeomorphic representations for the groupwise analysis of non-rigid registrations of medical images. T.M.I. (2004) 1006–1020
4. Zollei, L., Learned-Miller, E., Grimson, E., Wells, W.: Efficient population registration of 3d data. In: C.V.B.I.A. Volume 3765 of LNCS., Springer (2005) 291–301
5. Glaunès, J., Joshi, S.: Template estimation from unlabeled point set data and surfaces for computational anatomy. In Pennec, X., Joshi, S., eds.: Proc. Workshop on the Mathematical Foundations of Computational Anatomy. (2006)
6. Chui, H., Rangarajan, A., Zhang, J., Leonard, C.: Unsupervised learning of an atlas from unlabeled point-sets. Trans. on Patt. Anal. and Machine Intell. **26**(2) (2004) 160–172
7. Allassonnière, S., Amit, Y., Trouvé, A.: Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society Series B **69**(1) (2007) 3–29
8. Ma, J., Miller, M., Trouvé, A., Younes, L.: Bayesian template estimation in computational anatomy. To appear in NeuroImage (2008)
9. Allassonnière, S., Kuhn, E., Trouvé, A.: Bayesian deformable models building via a stochastic approximation algorithm: A convergence study. Bernouilli Journal (2008) In revision.
10. anonymous.
11. Vaillant, M., Glaunès, J.: Surface matching via currents. In: Proceedings of Information Processing in Medical Imaging (2005), LNCS 3565. 381–392
12. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Measuring brain variability via sulcal lines registration: a diffeomorphic approach. In: Proc. MICCAI, LNCS 4791. (2007) 675–682
13. Glaunès, J.: Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique. PhD thesis, Université Paris 13, http://cis.jhu.edu/ joan/TheseGlaunes.pdf (2005)
14. Aronszajn, N.: Theory of reproducing kernels. Trans. Amer. Math. Soc. (68) (1950)
15. Trouvé, A.: Diffeomorphisms groups and pattern matching in image analysis. International Journal of Computer Vision **28** (1998) 213–221
16. Dupuis, P., Grenander, U., Miller, M.: Variational problems on flows of diffeomorphisms for image matching. Quaterly of Applied Mathematics (1998)
17. Pati, Y., R.Rezaifar, Krishnaprasad, P.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: 27th Asilomar Conf. on Signals, Systems and Comput. (November 1993)
18. Fillard, P., al.: Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. NeuroImage **34**(2) (January 2007) 639–650
19. Hazlett, H., Poe, M., Gerig, G., Smith, R., Provenzale, J., Ross, A., Gilmore, J., Piven, J.: Magnetic resonance imaging and head circumference study of brain size in autism. The Archives of General Psychiatry **62** (2005) 1366–1376