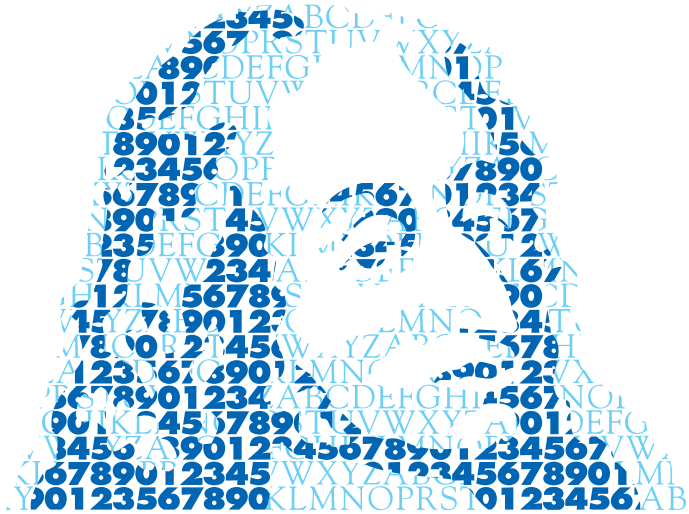
ANNALES MATHÉMATIQUES



BLAISE PASCAL

Stéphanie Allassonnière, Jérémie Bigot, Joan Alexis Glaunès, Florian Maire & Frédéric J.P. Richard

Statistical models for deformable templates in image and shape analysis

Volume 20, n° 1 (2013), p. 1-35.

<http://ambp.cedram.org/item?id=AMBP_2013__20_1_1_0>

Publication éditée par le laboratoire de mathématiques de l'université Blaise-Pascal, UMR 6620 du CNRS Clermont-Ferrand — France

cedram

Article mis en ligne dans le cadre du Centre de diffusion des revues académiques de mathématiques
http://www.cedram.org/

Annales mathématiques Blaise Pascal **20**, 1-35 (2013)

Statistical models for deformable templates in image and shape analysis

Stéphanie Allassonnière
Jérémie Bigot
Joan Alexis Glaunès
Florian Maire
Frédéric J.P. Richard

Abstract

High dimensional data are more and more frequent in many application fields. It becomes particularly important to be able to extract meaningful features from these data sets. Deformable template model is a popular way to achieve this. This paper is a review on the statistical aspects of this model as well as its generalizations. We describe the different mathematical frameworks to handle different data types as well as the deformations. We recall the theoretical convergence properties of the estimators and the numerical algorithm to achieve them. We end with some published examples.

Modèles statistiques d'atlas déformables pour l'analyse d'images et de formes

Résumé

Les données de grande dimensions sont de plus en plus fréquemment collectées dans de nombreux domaines d'application. Il devient alors particulièrement important d'être capable d'extraire des caractéristiques significatives de ces bases de données. Le modèle d'atlas déformable (Deformable template model) est un outil maintenant répandu pour atteindre ce but. Cet article présente un panorama des aspects statistiques de ce modèle ainsi que ses généralisations. Nous décrivons les différents cadres mathématiques permettant de prendre en compte des types variés de données et de déformations. Nous rappelons les propriétés théoriques de convergence des estimateurs et des algorithmes permettant l'estimation de ces caractéristiques. Nous terminons cet article par la présentation de quelques résultats publiés utilisant des données réelles.

Keywords: Review paper, Deformable template model, statistical analysis.
Math. classification: 62H12, 62H30, 62H35.

Contents

1. Introduction 2
2. An overview of deformable template models 4
2.1. Types of data 4
2.2. Various deformation models 7
2.3. Mixture models 10
2.4. Spatiotemporal models 11
3. Consistent estimation of a template 13
3.1. Nonparametric approach 13
3.2. Parametric models of non-rigid deformations 15
3.2.1. A frequentist approach using large deformations 15
3.2.2. A Bayesian approach using small deformations 16
4. Algorithms for atlas estimation 18
4.1. Deterministic approach 18
4.2. Stochastic approach 19
5. Applications 22
5.1. Estimated templates 22
5.2. Clustering and Classification 23
5.2.1. Classification via maximum a posteriori 23
5.2.2. Classifying registration 25
5.3. Spatiotemporal templates 26
6. Conclusion 27
References 29

1. **Introduction**

Due to the increasing number of images available in many fields, image analysis has become an important challenge. This is particularly the case in the medical domain where images of different modalities are easier and easier to acquire. This amount of data paves the way to some population analysis, characterizing the population in terms of representative features, clustering the heterogeneous populations and of course classifying new observations. This review paper focuses on methods providing a statistical analysis of population in terms of both estimating a representative mean image and some characteristic geometrical variations within the observed population.

We are interested in the study of statistical deformable models whose ideas go back to D'Arcy Thompson in his book "On Growth and form" [49] for the study of animal and human anatomical shapes and which were generalized and formulated mathematically by Grenander in [31] for image analysis. The key idea is to assume that, for a given homogeneous population, each observation is a regular deformation of a mean representative image, called template, plus some additive noise. The deformation does not affect the template itself but moves the underlying space the image is painted on, carrying the color information from one location to another. The deformations are also characteristic of the population, making a difference among populations in terms of both the photometry (through the template) and the geometry (through the deformation distribution). They together will be called the atlas in the sequel. The main issue is to learn this atlas as the parameter of a statistical generative model when we observe a population of images. We will summarize here the different aspects of this model and the theoretical guaranties we have for this statistical estimation problem.

The general framework in this setting is the following: let $\mathcal{H}$ be a separable Hilbert space modeling a set of objects and $\mathcal{G}$ be a group acting on $\mathcal{H}$, meaning that for any $(h, \phi) \in \mathcal{H} \times \mathcal{G}$ the (left) action $\cdot$ of $\mathcal{G}$ onto $\mathcal{H}$ is such that $\phi \cdot h \in \mathcal{H}$. Suppose that the data at hand $Y_1, \ldots, Y_n$ are i.i.d random variables in $\mathcal{H}$ satisfying the model:

$$Y_i = \phi_i \cdot \mathcal{T} + W_i, \tag{1.1}$$

where $\mathcal{T} \in \mathcal{H}$ is an unknown template, $W_i \in \mathcal{H}$ is an additive noise (typically due to the measurement process) independent of the $\{\phi_i\}_{1 \leq i \leq n}$, which are i.i.d random variables belonging to $\mathcal{G}$ modeling a geometric source of variability in the data. Given a template, the specification of a geometric deformation and an additive noise model, observations $\{Y_n\}_{n>0}$ can be simulated as independent realizations of the generative stochastic process (1.1).

We present successively in this paper some classes of objects handled by this model and some typical deformation groups. These will specify the mathematical frameworks which have to be used to deal with the large variety of observations we encounter. Some extensions of model (1.1), which include time-dependency as well as mixture models, and enable population clustering, are also described to overcome some model limitations. In

Section 3, we outline the theoretical consistency of the template estimator and full atlas estimation issues in some of the presented mathematical frameworks. Section 4 presents the most efficient and popular algorithms to numerically perform this estimation. The theoretical convergence of these algorithms are recalled. The paper ends with presenting some results, obtained by applying these methods to several types of data sets.

2. An overview of deformable template models

The deformable template model (1.1) applies to many different situations. In this section, we give an overview of the different aspects of the same model focusing on the data and deformations.

2.1. Types of data

The typical data considered by the deformable template model are signals, functions from $\Omega \subset \mathbb{R}$ to $\mathbb{R}$, or images of dimension d, functions from $\Omega \subset \mathbb{R}^d$ to $\mathbb{R}$, where Ω is a compact subset and $d = 2, 3$ or 4, including time sequences. In this setting, the Hilbert space $\mathcal{H}$ is the space of squared integrable functions $\mathcal{H} = L^2(\Omega, dx)$ with respect to the Lebesgue measure. To define the gray level of the template $\mathcal{T}$ on a given point $x \in \Omega$, it is required to have more regularity on this template function. It is usually assumed that the template is at least piecewise continuous.

When we consider real data, we are usually provided with digital inputs which are discrete, giving gray level value on a regular grid of points. To adapt the deformable template model (1.1), we assume that $\mathcal{T}$ is defined on the whole subset Ω and belongs to $\mathcal{H}$ whereas each observation Y_i is a discretization of the deformed template on a regular grid. This dimension reduction step is also shared by the noise structure W which therefore can be defined as a random vector of finite dimension. This discretization implies that the observations do not actually belong to the same Hilbert space as the template but for sake of simplicity we omit these two subsamplings for Y_i and W_i in the notation of Eq. (1.1).

Concerning the action of a deformation onto such a template, the idea is that the image (grey level map on Ω) is carried by the underlying space Ω of $\mathbb{R}^d$ defined by its coordinate system. The deformation moves the points in the space which transports the grey level information. Therefore, the grey level at location x of the deformed template is the grey level of the

template at location $\phi^{-1}(x)$ when ϕ is invertible. This writes :

$$\phi \cdot \mathcal{T}(x) = \mathcal{T}(\phi^{-1}(x)), \quad \forall x \in \Omega\,. \tag{2.1}$$

For the definition of this action, two conditions are required. First, we need to consider invertible ϕ's or to provide a meaningful approximation of ϕ^{-1}. Moreover, the deformation ϕ should be defined on Ω.

In many applications it is preferable to extract meaningful features from the image data prior to perform shape analysis. Considering the case of medical image data, such features will usually be derived from manual or automated segmentations of anatomical landmarks on the images. From a mathematical modeling viewpoint, such features take the form of sparse geometrical data lying in the ambient space: points or groups of points, curves or surfaces, needing specific models to be treated. More precisely, if we want to stick to our template/deformation model, we need to define specific Hilbert spaces for each kind of geometrical data in use. The first and easier type is the so-called *landmarks case*, where features form a sequence of n labeled points in space Ω. In this case we set $\mathcal{H}$ to be the Euclidean space $(\mathbb{R}^d)^n$, and the action of any deformation map $\phi \in \mathcal{G}$ on $x = (x_1, \ldots, x_n) \in \mathcal{H}$ will simply write $\phi.x = (\phi(x_1), \ldots, \phi(x_n))$.

When it comes to unlabeled groups of points, the previous model is not suitable. It has been proposed in [30] to model such features as sums of Dirac functionals embedded in the dual space of a functional Hilbert space. More precisely, an unlabeled set of points $\{x_1, \ldots, x_n\} \subset \Omega$ is modeled as the measure $\mu = \sum_{i=1}^n a_i \delta_{x_i}$, where non-equal weights a_i may model differences in relevancy or confidence of locations between subsets of the set of points. Next, assume there exists a Hilbert space H of functions defined on Ω such that all evaluations functionals $\delta_x : f \mapsto f(x),\ x \in \Omega$ are continuous linear forms on H, or in other words $\mathrm{Vect}\{\delta_x, x \in \Omega\} \subset H^*$. We set $\mathcal{H} = H^*$, the dual space of H, together with its dual Hilbert metric: for any $\mu \in \mathcal{H}$, $\|\mu\|_{\mathcal{H}} = \mathrm{Sup}\{\mu(f),\ f \in H,\ \|f\|_H \leq 1\}$. For the objects in use, sums of Dirac deltas, this dual metric takes an explicit form which gives to this model one of its main practical interest: if $\mu = \sum_{i=1}^n a_i \delta_{x_i}$, then

$$\|\mu\|^2_{\mathcal{H}} = \sum_{i=1}^n \sum_{j=1}^n a_i a_j k_H(x_i, x_j),$$

where k_H denotes the reproducing kernel of space H. Indeed the assumption made on H exactly states that H is a Reproducing Kernel Hilbert Space (RKHS), and the space H is uniquely specified by its kernel k_H. The action of a deformation map $\phi \in \mathcal{G}$ on an element $\mu \in \mathcal{H}$ is the usual push-forward of the measure μ, defined by duality as $\forall f \in H, (\phi.\mu)(f) = \mu(\phi.f) = \mu(f \circ \phi)$. For sums of Dirac deltas, we get simply the explicit formula: if $\mu = \sum_{i=1}^n a_i \delta_{x_i}$, then $\phi.\mu = \sum_{i=1}^n a_i \delta_{\phi(x_i)}$.

The previous measure framework can be used directly to treat the case of curves or surfaces embedded in Ω, because such objects can be modeled as measures. For example let $\mathcal{C} \subset \Omega$ be a curve, parameterized by $\gamma : [0,1] \to \Omega$. Then, one can associate to $\mathcal{C}$ a measure $\mu \in \mathcal{H}$ via the rule

$$\forall f \in H, \quad \mu(f) = \int_0^1 f(\gamma(s))\|\gamma'(s)\|ds.$$

However this does not provide a purely geometric model, since the action of deformation maps will in general modify the mass repartition on the image curve. In other words, the measure $\phi.\mu$ will not be in general the uniform measure associated to the curve $\phi(\mathcal{C})$. A purely geometrical model is obtained when one replaces measures by currents, i.e. measures acting on spaces of differential forms. In this new setting, the space H is replaced by a Hilbert space W of m-differential forms, where m is the dimensionality of the data; and the current $\vec{\mu}$ associated to an oriented m-submanifold S of Ω is defined via the rule $\vec{\mu}(\omega) = \int_S \omega$. In the example of curve $\mathcal{C}$, this writes $\forall \omega \in W, \quad \vec{\mu}(\omega) = \int_0^1 \omega(\gamma(s))(\gamma'(s))ds$, and the dual norm in $\mathcal{H} = W^*$ writes in terms of the kernel k_W, which in its general form is a map from Ω^2 to the space of symmetric bilinear forms of $\mathbb{R}^d$:

$$\|\vec{\mu}\|^2_{\mathcal{H}} = \int_0^1\int_0^1 k_W(\gamma(s),\gamma(t))(\gamma'(s),\gamma'(t))ds\ dt.$$

The push-forward of a deformation map $\phi \in \mathcal{G}$ on an element $\vec{\mu} \in \mathcal{H}$ is defined via duality as $\forall \omega \in W, \quad (\phi.\vec{\mu})(\omega) = \vec{\mu}(\phi.\omega)$, where for all $x \in \Omega$ and $\alpha_1, \ldots, \alpha_m \in \mathbb{R}^d$,

$$(\phi.\omega)(x)(\alpha_1, \ldots, \alpha_m) = \omega(\phi(x))(d\phi(x).\alpha_1, \ldots, d\phi(x).\alpha_m).$$

As opposed to the scalar case, this action is consistent with the action of ϕ on subsets: if $\vec{\mu}$ is the current associated to an oriented m-submanifold S, then $\phi.\vec{\mu}$ will be the current associated to $\phi(S)$.

We will not detail any further this specific representation and refer to [53, 29] for developments for the case of curves and surfaces.

2.2. **Various deformation models**

A simple setting to study the problem of estimating the template $\mathcal{T}$ in model (1.1) is to consider the case where $\mathcal{G}$ is a connected, semi-simple, and compact Lie group, and $\mathcal{H} = L^2(\mathcal{G})$ the Hilbert space of complex valued, square integrable functions on the group $\mathcal{G}$ with respect to the Haar measure $\mathrm{d}\phi$. In [11] and [12], it has been proposed to consider the nonparametric estimation of a complex valued template $\mathcal{T} : \mathcal{G} \to \mathbb{C}$ in the following deformable white noise model:

$$\mathrm{d}Y_i(\phi) \;=\; \mathcal{T}(\phi_i^{-1} \cdot \phi)\,\mathrm{d}\phi + \varepsilon\,\mathrm{d}W_i(\phi), \quad \phi \in \mathcal{G},\ i \in [\![1, n]\!] \qquad (2.2)$$

where the W_i's are independent copies of a standard Brownian sheet W on the topological space $\mathcal{G}$ with reference measure $\mathrm{d}\phi$, $\epsilon > 0$ is the level of noise in the measurements, and the random variables ϕ_i are supposed to be i.i.d. with a known density $h \in L^2(\mathcal{G})$. The white noise model (2.2) is a continuous model that is useful for studying the statistical estimation of $\mathcal{T}$ in the asymptotic setting $n \to +\infty$ as it avoids the problem of controlling the bias introduced by any discretization scheme. Some concrete examples of model (2.2) include the analysis of translated two-dimensional images, which corresponds to the case $\mathcal{G} = \mathbb{R}^2/\mathbb{Z}^2$ (the torus in dimension two), and which founds its applications in biomedical imaging or satellite remote sensing (see e.g. [26]). Other examples are rotation models for two-dimensional or three-dimensional images for which either $\mathcal{G} = \mathrm{SO}(2)$ or $\mathcal{G} = \mathrm{SO}(3)$ (the special orthogonal group in dimension 2 or 3) when the images at hand are observed through the action of random rotations (see e.g. [32, 42, 45]). Note that model (2.2) corresponds to a specific case where the group $\mathcal{G}$ and the domain Ω (on which the template $\mathcal{T}$ is defined) are the same. In practice the data are rather observed on a discrete grid of time points or pixels belonging to a convex set $\Omega \subset \mathbb{R}^d$. Thus, an alternative model is

$$Y_{i,\ell} \;=\; \mathcal{T}(\phi_i^{-1} \cdot x_\ell) + \sigma W_{i,\ell}, \quad x_\ell \in \Omega,\ i \in [\![1, n]\!],\ \ell \in [\![1, p]\!], \qquad (2.3)$$

where $\mathcal{T} \in \mathcal{H} = L^2(\Omega)$, the ϕ_i's are i.i.d. random variables in $\mathcal{G}$, the $W_{i,\ell}$ are i.i.d. standard Gaussian variables, $\sigma > 0$ is the level of additive noise, and the x_ℓ's are a set of p deterministic sampling points in Ω (e.g. $x_\ell = \ell/p$ in the setting $\Omega = [0, 1]$). Model (2.3) assumes that $\mathcal{G}$ is a group acting on the domain Ω meaning $\phi \cdot x \in \Omega$ for any $(\phi, x) \in \mathcal{G} \times \Omega$. In model (2.3), the main goal is to construct consistent estimators of the template $\mathcal{T}$ and to recover the geometric variability of the data modeled by the ϕ_i's in the

double asymptotic setting $\min(n,p) \to +\infty$ (both the number of images and the number of sampling points tend to infinity).

The compact groups reduce to deformations which are highly constrained, typically rigid body deformations. Therefore, to include deformations with more degree of freedom, one can think of elastic maps. A first step in this generalization considers that a deformation is given by the displacement of each point in the definition domain Ω. This yields the existence of a velocity vector field, denoted $v : \Omega \to \Omega$, such that for all $x \in \Omega$, $\phi(x) = x + v(x)$. Note that there is no constraint on v which ensures the invertibility of ϕ. This requires to provide an adapted definition of the previously introduced actions when it involves the inverse of ϕ. The inverse of the mapping is approximated at first order ($\|v\|_{\mathcal{G}}$ small) by $Id - v$, where Id is the identity map on $\mathbb{R}^d$.

This framework has a big advantage as it provides an easy characterization of the deformation through v. However, this remains an infinite dimensional variable to consider and estimate. To reduce this problem to a finite dimensional one, one may assume that this vector has a finite decomposition on a dictionary basis. The velocity field is stated as follows. Given a dictionary of functions $(b_k)_{1\leq k\leq k_{\mathcal{G}}}$ from Ω to Ω, for all $\phi \in \mathcal{G}$, there exists $(\alpha_k)_{1\leq k\leq k_{\mathcal{G}}}$ such that for all $x \in \Omega$

$$\phi(x) = x + \sum_{k=1}^{k_{\mathcal{G}}} \alpha_k b_k(x). \tag{2.4}$$

The regularity of ϕ depends on the choice of the dictionary basis. This parametric model aims at constructing a consistent estimator of the full atlas as the number of observations tends to infinity.

As already pointed, the previous model of elastic, "small" deformations does not guarantee invertibility of the maps. A more rigorous approach takes the previous framework as a model for infinitesimal deformations only, building a group of diffeomorphisms by integrating such displacement fields. This construction has been introduced in several works [50, 18] and has been intensely developed since then to derive efficient techniques for shape registration and analysis (see e.g. [36, 44, 9, 27, 6], among many others). The construction starts by specifying a Hilbert norm on the space V of displacement fields (see [50]). Then we consider time-dependent families of such vector-fields $v \in L^2([0,1], V)$. The deformation map ϕ^v associated

to v is defined as the flow at time one of this family: $\phi^v = \phi_1$, where ϕ_t satisfies the O.D.E.

$$\begin{cases} \frac{\partial \phi_t(x)}{\partial t} = v_t(\phi_t(x)), \\ \phi_0(x) = x. \end{cases} \tag{2.5}$$

The induced group of deformation maps $\mathcal{G} := \{\phi^v, v \in V\}$ is provided a right-invariant metric via the rule $d(Id, \phi) := \sqrt{\inf\{\int_0^1 \|v_t\|_V^2 dt, \phi = \phi^v\}}$. This exactly states that $\mathcal{G}$ is given the structure of an infinite-dimensional manifold, on which distances are computed as the length of minimal geodesic paths connecting two elements. In fact it is possible to derive the equations of geodesics, which can be seen as a special case of the Euler-Poincaré equations of motion, coined as EPDiff ([35]). This allows to consider a new parameterization of deformation maps $\phi \in \mathcal{G}$ via the initial velocity field v_0. For simplicity, we only detail the finite-dimensional case below.

The large deformation representation boils down to a finite dimensional one when carried out in a discrete setting. Indeed, when the support of the data in use forms a finite set of points (as in the case of landmarks or point clouds), or can be approximated with a finite set of points (as e.g. for curves, surfaces or signals/images), then one only needs to consider initial velocity maps which are linear combinations of kernel functions centered at these points. More explicitly, if this set of discretization points are $(x_i)_{1 \leq i \leq n}$, then the initial velocity map can be written as

$$v_0(x) = \sum_{i=1}^{n} k_V(x_i, x)\alpha_i,$$

where the $\alpha_i \in \mathbb{R}^d$ are the coefficients of the linear combination and are called *momentum vectors*. Equivalently it is possible to parameterize v_0 via the velocity vectors $v_0(x_i)$. The complete system of ODEs that express the deformation map ϕ parameterized by a finite set of initial momentum vectors α_i is the following:

$$\begin{cases} \frac{dx_j(t)}{dt} = \sum_{i=1}^{n} k_V(x_i(t), x_j(t))\alpha_i(t) \\ \frac{d\alpha_j(t)}{dt} = -\frac{1}{2}\sum_{i=1}^{n} \frac{\partial}{\partial x_i(t)} \left[k_V(x_i(t), x_j(t))\alpha_i(t)\right] \alpha_j(t) \\ \frac{d\phi_t(x)}{dt} = \sum_{i=1}^{n} k_V(x_i(t), \phi_t(x))\alpha_i(t), \end{cases}$$

with $x_i(0) = x_i$, $\alpha_i(0) = \alpha_i$, $\phi_0(x) = x$. The first two equations form the geodesic equations, while the third one is Eq. 2.5 in this discrete setting.

Note that these geodesic equations in this finite dimensional case enable to parameterize the deformation by the initial momenta as well as the initial control point locations.

2.3. **Mixture models**

In the observation model (2.2), additive random fields W_i are critical for the modeling of statistical dependencies between values of the deformed template $\phi_i \cdot \mathcal{T}$ and observations Y_i. For instance, when objects are images, they describe how intensities (gray-level values) of observations relates to the template ones. Assumptions made on the probability laws of these fields completely determine the nature of dependencies between observations and template. In particular, the basic assumption that fields W_i are white noises implies that dependencies have the same statistical properties all over the space.

There are many situations where such an assumption is not valid (see for instance [34] and Section 5.2). For instance, a problem arises in contrast-enhanced medical imaging when two tissues having different contrast-agent absorption properties are observed with a same intensity range before contrast injection but two different ones after injection. In such a case, relationships between intensities of pre and post contrast images depend on tissue types, and are not spatially invariant.

To deal with such a situation, [34] proposed to define the probability distribution of W_i conditionally to the class of the pixel. Let us assume that pixels of the domain Ω can be divided into K classes where template-observation dependencies are different, and denote by $L_k(x)$ the probability for a pixel x to belong to the kth class. The probability distribution of $W_i(x)$ at x is then defined conditionally to $L(x) = (L_1(x), \cdots, L_K(x))$ as a mixture of distributions π_k describing dependencies on each class k :

$$\pi(W_i(x) = w_x | L(x)) = \sum_{k=1}^{K} L_k(x) \pi_k(W_i(x) = w_x), \ \forall\, w_x \in \mathbb{R}.$$

In the defined model, the so-called class map $\mathbf{L} = \{L(x_\ell), x_\ell \in \Omega, \ell \in [\![1, p]\!]\}$ is unknown and has to be estimated together with deformations. In [34], the estimation of deformations and class map is jointly performed by a MAP approach. It leads to a simultaneous classification of template

pixels and registrations of the template and observations, which is called classifying registration.

Besides, some spatial constraints on pixel classes can be added to the model by specifying a probability distribution on the map $\mathbf{L}$. For instance, in the case when $K = 2$ (two classes), we can set that we restrict the amount of pixels of the class 0, whereas we enforce the spatial homogeneity of the classes.

A mixture of the deformable template model (1.1) involving several templates can also be considered. It is particularly appealing when one wants to extract reference curves or patterns from data featuring high dispersion [3, 41]. In this setting, the data are described with a set of K templates $\{\mathcal{T}_1, \ldots, \mathcal{T}_K\}$ and each observation Y_i is assigned to a (possibly unknown) class $J_i \in \{1, \ldots, K\}$ such that given $J_i = j$, (1.1) writes:

$$Y_i = \phi_i \cdot \mathcal{T}_j + W_i \,. \tag{2.6}$$

In a parametric approach, each elementary component is characterized by its own parameter set θ_j. In particular, each has its own template function $\mathcal{T}_j$, a prior deformation distribution p_j and a prior weight w_j, such that $\sum_{j=1}^{K} w_j = 1$. As a result, the mixture of the deformable model can be fully parameterized by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$. The number of classes is fixed but priors on the parameters or other arguments enables to achieve an estimation of the optimal number for the analyzed population.

Retrieving the data diversity becomes possible by using a mixture model to estimate the templates $\{\mathcal{T}_1, \ldots, \mathcal{T}_K\}$, which can then be used to perform unsupervised classification (see section 5).

2.4. **Spatiotemporal models**

In some cases, observations in model (2.2) are sequential, and can be naturally ordered by time. This is for instance the case in dynamic contrast enhanced imaging where a series of images of a same body part is produced to capture the dynamic of contrast agent absorption by tissues (see [33] and Section 5.3). In such a situation, time variations of observations may not be exclusively geometrical (*i.e.* caused by deformations). Variations may also appear as changes in observation values (photometric changes when observations are images). In dynamic contrast enhanced imaging, these variations are due to the contrast agent diffusion. This other source of variations can be included in the model by defining a time-dependent

template. Letting $t_1 < \cdots < t_m$ be acquisition times, the observation model is then of the form:

$$Y_{(i)} = \phi_{(i)} \cdot \mathcal{T}_{(i)} + W_{(i)}, \tag{2.7}$$

where $Y_{(i)}$, $\phi_{(i)}$, $\mathcal{T}_{(i)}$, $W_{(i)}$ are the observation, the deformation, the template and the noise at time t_i, respectively.

However, the model defined above cannot be used without setting some constraints on the temporal template. Indeed, defined with a large degree of freedom, the temporal template can not only account for photometric variations but also for geometrical variations. In other words, without limiting the degree of freedom, the two sources of observation variations are undistinguishable in the model. One way of fixing this issue consists in restricting the template definition to a class of parametric functions which is representative of the expected photometric variations. Such a solution was proposed in [33] in the context of dynamic contrast enhanced imaging. The template was defined using a class of functions which are known to be well-suited for fitting dynamic curves of contrast absorption. This class is characterized by just a few interpretable parameters.

In [33], the model is further constrained by partitionning pixels of Ω into classes of common temporal photometric variation: for all pixels x belonging to a same class k, $\mathcal{T}_{(i)}(x) = f_k(t_i), i = 1, \cdots, m$, where f_k is a parametric temporal curve for the kth class. Pixel classes and associated temporal curves are both estimated at the same time using an iterative algorithm, while the number of classes is set automatically during iterations. In this model, photometric variations are estimated globally on pixels of a same class rather than locally on each pixel. This reduces the influence of local geometrical variations on the estimation of photometric variations.

Another aspect of the dynamical model is to consider that the time dependent template is a smooth deformation of a baseline at a reference time point. This is particularly adapted to the evolution of shapes of brain structures such as hippocampus, caudate or basal ganglia along development or aging. Contrary to the spatiotemporal template previously described, it does not enable a change in contrast or equivalently topology of the shape. However, this framework enables to capture a subject growth model and also to produce a global template which evolves in time and space. This provides a mean growth scenario of given time evolving population.

Constructing a growth model of a single subject is similar to perform regression on a set of images. When a sequence of images is observed for a subject, one can estimate the evolution scenario from the first time point to the last observed one by deforming the initial shape so that it matches the intermediate observations. Several models have been proposed which differ from each other in the smoothness of the time dependency of the regression model ([22, 55, 23]). In any case, the model writes :

$$\phi_{(t)}(\mathcal{T}) \triangleq \mathcal{T}_{(t)} \quad \text{such that} \quad \mathcal{T}_{(t_i)} \sim Y_{(i)}\,, \tag{2.8}$$

where $(Y_{(i)})_{1\leq i\leq p}$ is the longitudinal data, $\mathcal{T}$ is the baseline image (that could be fixed as the first acquired image or estimated in a joint optimization) and $\phi_{(t)}$ is the continuous regression function.

This scheme has been extended to estimate a spatiotemporal atlas from longitudinal data in which different subjects have been observed at several time points. The spatiotemporal atlas consists of a mean image, its evolution along time, together called the mean scenario of the population, as well as the correspondences that map this mean scenario to evolutions of n subjects. There are different ways to transport the mean scenario (time varying flow of deformation) from the template space to the space of the subject. In [46, 40, 39], the mapping is done by parallel transport. In [22], an adjoint transport has been proposed together with possible change in time dynamics. The observations are surfaces, the template is therefore a current and the deformation is assumed diffeomorphic through the LDDMM framework introduced above. We refer to [22] for all the details of this atlas estimation. Note that this framework applies directly with any other type of data such as images.

3. **Consistent estimation of a template**

We summarize here the theoretical results of asymptotic convergence for the template estimation problem. Considering different situations, we describe the statistical consistency properties which can be achieved.

3.1. **Nonparametric approach**

A nonparametric approach is to consider that the template $\mathcal{T}$ belongs to an infinite dimensional space $\mathcal{F} \subset \mathcal{H}$ such as a Sobolev ball, and that it cannot be parameterized by a finite number of parameters. In [11] and

[12], it is shown that, in model (2.2), the density h of the random elements $\phi_i \in \mathcal{G}$ plays the role of the kernel a convolution operator that has to be inverted to construct an optimal estimator of $\mathcal{T}$. Indeed, in (2.2), $\mathbb{E}\mathcal{T}(\phi_i^{-1}\phi) = \int_{\mathcal{G}} \mathcal{T}(\tilde{\phi}^{-1}\phi)h(\tilde{\phi})\,\mathrm{d}\tilde{\phi}$ is the convolution over the group $\mathcal{G}$ between the template $\mathcal{T}$ and the density h. Hence, it is possible to build an estimator of $\mathcal{T}$ using a regularized deconvolution method over Lie groups. This class of inverse problems is based on the use of harmonic analysis and Fourier analysis on compact Lie groups to transform convolution in a product of Fourier coefficients. In [11] and [12], it is shown that such estimators of $\mathcal{T}$ achieve an optimal rate of convergence over Sobolev balls in an asymptotic setting, where the number n of images tends to infinity. A more standard approach to estimate a template is a two step procedure which consists in first computing estimators $\hat{\phi}_1, \ldots, \hat{\phi}_n$ of the unobserved deformations $\phi_1, \ldots, \phi_n$ and then in averaging the data after an alignment step which yields to an estimator of the following form

$$\hat{\mathcal{T}}_0(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{\mathcal{T}}_i(\hat{\phi}_i \cdot x),\ x \in \Omega, \tag{3.1}$$

where $\hat{\mathcal{T}}_i(x)$ denotes some estimator of the unknown i-th image $\mathcal{T}(\phi_i^{-1}x)$ in either model (2.2) or model (2.3). Such a procedure clearly depends on the quality of the estimation of the unobserved deformations. In the simple setting where $\Omega = \mathcal{G} = \mathbb{S}_1 \simeq [0,1[$ (the torus in dimension 1), the following lower bounds have been established in [10] and [12]: for *any* estimators $(\hat{\phi}_1, \ldots, \hat{\phi}_n) \in \mathbb{S}_1^n$ and under mild assumptions on $\mathcal{T}$ and the density h of the random variables $\phi_1, \ldots, \phi_n$, one has that in model (2.2)

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{\phi}_i - \phi_i)^2\right) \geq \frac{\epsilon^2}{\int_0^1 \left|\frac{\partial}{\partial x}\mathcal{T}(x)\right|^2 dx + \epsilon^2\mathcal{I}(h)} \underset{n\to+\infty}{\nrightarrow} 0 \tag{3.2}$$

and in model (2.3)

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{\phi}_i - \phi_i)^2\right) \geq \frac{\sigma^2}{p\left(\int_0^1 \left|\frac{\partial}{\partial x}\mathcal{T}(x)\right|^2 dx\right) + \sigma^2\mathcal{I}(h)} \underset{p\to+\infty}{\to} 0 \tag{3.3}$$

where $\mathcal{I}(h) = \int \left(\frac{\partial}{\partial x}\log h(x)\right)^2 h(x)dx$. Therefore, Eq. (3.2) shows that, in model (2.2) and in the asymptotic setting $n \to +\infty$, it is not possible to build consistent estimators of the deformations in the sense that

$\liminf_{n\to+\infty} \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{\phi}_i - \phi_i)^2\right) \neq 0$. Thus, in such a setting, estimators $\hat{\mathcal{T}}_0$ based on an alignment step (as defined in (3.1)) are not likely to be consistent estimators of the template $\mathcal{T}$. To the contrary, in model (2.3) and in the double asymptotic setting $\min(n,p) \to +\infty$, equation (3.3) suggests that one can compute consistent estimators of the deformations parameters. This implies that, in model (2.3), it is possible to construct estimators $\hat{\mathcal{T}}_0$ via an alignment procedure such that

$$\lim_{\min(n,p)\to+\infty} \left\|\hat{\mathcal{T}}_0 - \mathcal{T}\right\|_{\mathcal{H}}^2 = 0.$$

For a detailed discussion, we refer to [10].

To conclude this section, we would like to mention that another point of view for statistical inference in either model (2.2) or (2.3), is to consider that the deformations $\phi_i, i = 1, \ldots, n$ are fixed (non-random) parameters belonging a finite dimensional Lie group. For various results in this semi-parametric framework, we refer to [13, 14, 25, 56].

3.2. **Parametric models of non-rigid deformations**

3.2.1. **A frequentist approach using large deformations**

Now consider the following case. The data are 2D images observed on a grid of pixels $x_\ell \in \Omega,\ \ell = 1, \ldots, p$, and the deformations in model (1.1) are large random diffeomorphisms ϕ_a of Ω that are solutions of an O.D.E governed by vector fields $v_a : \mathbb{R}^2 \to \mathbb{R}^2$ parametrized by a set of coefficients $a \in [-A, A]^{2K}$ for a given real $A > 0$, see [15] for further details.

To generate a large class of random diffeomorphisms, it is proposed in [15] to consider a random distribution on the vector a. Let us denote by $\mathbb{P}_A$ a probability distribution on $[-A, A]^{2K}$. The data at hand are then supposed to satisfy the model

$$Y_{i,\ell} = \mathcal{T}(\phi_{a_i}^{-1}(x_\ell)) + \sigma W_{i,\ell}, \quad x_\ell \in \Omega,\ i \in [\![1, n]\!],\ \ell \in [\![1, p]\!], \tag{3.4}$$

where the a_i's are i.i.d. random variables sampled from the distribution $\mathbb{P}_A$, the $W_{i,\ell}$'s represent an additive noise independent from the vectors a_i, and $\mathcal{T} : \Omega \to \mathbb{R}$ is an unknown template.

In [15], it has been proposed to estimate $\mathcal{T}$ by a frequentist approach using M-estimation techniques (see e.g. [52] for an introduction). Denote by $\mathcal{Z} = \{Z : \Omega \to \mathbb{R}\}$ a set of images uniformly bounded (e.g. by the

maximum gray-level), and consider the generative model

$$Y_\ell = \mathcal{T}(\phi_a^{-1}(x_\ell)) + \sigma W_\ell, \quad x_\ell \in \Omega, \ell \in [\![1, p]\!], \tag{3.5}$$

with $a \sim \mathbb{P}_A$ and W_ℓ some additive noise. Let us define the function f as

$$f(a, W, Z) = \min_{u \in [-A,A]^{2K}} \sum_{\ell=1}^{p} \left(Y_\ell - Z(\Phi^1_{v_u}(x_\ell)) \right)^2, \tag{3.6}$$

where Z is a given image in $\mathcal{Z}$. Intuitively, f must be understood as the optimal cost to align the image Z onto the random image $Y = (Y_\ell)_{1 \leq \ell \leq p}$ using a deformation parametrized by some $u \in [-A, A]^{2K}$. At last, we define a contrast function F as

$$F(Z) = \int_{[-A;A]^{2K} \times \mathbb{R}^N} f(a, W, Z) d\mathbb{P}(a, W) = \mathbb{E} f(a, W, Z),$$

where $d\mathbb{P}(a, W)$ is the tensorial product measure on a. In practice, one can only compute the empirical contrast F_n as

$$\begin{aligned} F_n(Z) &= \int f(a, W, Z) d\mathbb{P}_n(a, W) \\ &= \frac{1}{n} \sum_{i=1}^{n} \min_{u_i \in [-A,A]^{2K}} \sum_{\ell=1}^{p} \left(Y_{i,\ell} - Z(\Phi^1_{v_{u_i}}(x_\ell)) \right)^2, \end{aligned}$$

where $\mathbb{P}_n(a, W) = \frac{1}{n} \sum_{i=1}^{n} \delta_{a^i, W_i}$. Then, one can define a sequence of sets of estimators as $\hat{Q}_n = \arg\min_{Z \in \mathcal{Z}} F_n(Z)$, whose asymptotic behavior is compared with the deterministic one $Q_0 = \arg\min_{Z \in \mathcal{Z}} F(Z)$. In [15], various conditions are discussed to ensure the consistency of the M-estimator $\hat{Q}_n$ in the sense that any accumulation point of a sequence of images $\hat{Z}_n \in \hat{Q}_n$ belongs to Q_0 almost surely.

3.2.2. **A Bayesian approach using small deformations**

We describe here the case where the template and the velocity vector fields of small deformations can be decomposed as a finite linear combination of some dictionary elements (see [1, 5]) as in Eq. (2.4). We focus on fixed sub-spaces determined by two sets of landmark points $(p_k)_{1 \leq k \leq k_\mathcal{H}}$ covering a larger domain $\Omega_p \supset \Omega$ and $(g_k)_{1 \leq k \leq k_\mathcal{G}} \in \Omega$. The template -resp. the deformation- is defined as a linear combination of a kernels $K_\mathcal{H}$ -resp. $K_\mathcal{G}$-

centered at the landmark points and are parameterized by the coefficients $w \in \mathbb{R}^{k_p}$ -resp. $\alpha \in \mathbb{R}^{d \times k_\mathcal{G}}$. We write

$$\mathcal{T}_w = \mathbf{K}_\mathcal{H} w, \text{ where } (\mathbf{K}_\mathcal{H} w)(x) = \sum_{k=1}^{k_\mathcal{H}} K_\mathcal{H}(x, p_k) w(k)\,, \tag{3.7}$$

$$\phi = Id + v, \text{ where } v_\alpha(x) = (\mathbf{K}_\mathcal{G}\alpha)(x) = \sum_{k=1}^{k_\mathcal{G}} K_\mathcal{G}(x, g_k)\alpha(k). \tag{3.8}$$

The model assumes that the observations are drawn with respect to the deformable template model (1.1). The deformations are however not observed and are treated as hidden random variables. This implies to state a distribution on the deformation coefficients $\alpha \in \mathbb{R}^{d \times k_\mathcal{G}}$ which are assume to follow a multivariate Gaussian distribution with zero mean and full covariance matrix $\Gamma_\mathcal{G}$. This has two advantages : first it introduces correlations between the movement of the control points. Moreover, it enables to catch the distribution of the observations as the marginal of the complete data $(Y_i, \phi_i)_{1 \le i \le n}$. This prevents from trying to estimate the best deformations which match the template to each observations Y_i which equals the mode of the posterior distributions of ϕ given $Y_i, \mathcal{T}$ (which is not consistent; c.f. (3.2) and (3.3)). This uncertainty is taken into account when marginalizing with respect to the deformation.

The complete statistical model writes : for all $1 \le i \le n$

$$\begin{cases} Y_i \sim \mathcal{T}(Id - \mathbf{K}_\mathcal{G}\alpha_i) + \sigma W_i \mid \alpha_i,\ \mathcal{T},\ \sigma \\ \alpha_i \sim \mathcal{N}(0, \Gamma_\mathcal{G}) \mid \Gamma_\mathcal{G} \end{cases} \tag{3.9}$$

where W_i are i.i.d standard normal random variables. Given this model, the parameters of interest, denoted by θ, are the template $\mathcal{T}$, the covariance matrix of the deformation space $\Gamma_\mathcal{G}$ and the noise variance σ^2.

The parameter estimates are obtained by maximizing the posterior likelihood on θ conditional on $(Y_i)_{1 \le i \le n}$ in a Bayesian framework. It has been proved in [1] that for any finite sample the maximum a posteriori exists and the MAP estimator is consistent. Without assuming that the observations are generated by the model described above but follow the distribution P we get : Let $\Theta_* = \{\ \theta_* \in \Theta \mid E_P(\log q(y|\theta_*)) = \sup_{\theta \in \Theta} E_P(\log q(y|\theta))\}$, where q is the observed likelihood.

Theorem 3.1 (Consistency on bounded prototypes)**.** *Under mild assumption on the model, we can prove that the restriction of* Θ_* *to bounded*

prototypes ($\|w\| \leq R$), Θ_*^R, is not empty and for any $\epsilon > 0$

$$\lim_{n\to\infty} P(\delta(\hat{\theta}_n^R, \Theta_*^R) \geq \epsilon) = 0\,,$$

where δ is any metric compatible with the topology on Θ^R.

4. **Algorithms for atlas estimation**

Now that we have the theoretical convergences ensured, the numerical schemes to compute these estimates are described.

4.1. **Deterministic approach**

The usual formulation of the template estimation is done by expressing the solution as the minimum of an energy. This energy is the sum of the contributions of the n observation registrations. Each of these contributions is a tradeoff between a fidelity to data term and a penalty on the deformation. This writes :

$$\hat{\mathcal{T}} = \operatorname*{argmin}_{\mathcal{T}\in\mathcal{H}} \min_{\phi_1,\dots,\phi_n} \sum_{i=1}^{n} \frac{1}{2\sigma^2}\|\phi_i \cdot \mathcal{T} - Y_i\|_{\mathcal{H}} + \frac{1}{2}\|\phi_i\|_{\mathcal{G}}\,, \tag{4.1}$$

where the tradeoff σ^2 has to be chosen by the user as a function of the details of the registration he/she expects.

The minimization of this energy is performed by optimizing iteratively with respect to the deformations and the template while fixing the other. Starting for an initial template (usually the mean of the gray level images), the n matching problems can be done separately as minimizing the n registration energies relates to the n subjects. This is computed by gradient descent most of the time due to the complex non linear dependency of the energy with respect to the deformation parameters (e.g. velocity coefficients of the linear combination of basis elements, initial momenta, etc). Given these best matches $(\hat{\phi}_i)_{1\leq i\leq n}$, the template is computed as the minimum of the sum of the n energies taken at $\phi_i = \hat{\phi}_i$. Many applications of this methods can be found in the literature for all the deformations and types of data presented above and many others [15, 48, 21, 37, 54].

This optimization scheme can be interpreted in two ways. The first one is a geometric interpretation. Looking for a template which satisfies Eq. (4.1) is similar to find the mean image whose orbit under the action of the group of deformations is the closest to the observations. This

approach is similar to find the Fréchet mean of the observations seen as random variables belonging to a nonlinear manifold. Indeed, the Fréchet mean [24] is an extension of the usual Euclidean mean to non-linear spaces endowed with non-Euclidean metrics. Minimizing the energy (4.1) can also be interpreted as finding a template which minimizes a dissimilarity measure that is a tradeoff between images registration and a penalty term to avoid too large deformations. This dissimilarity measure can be interpreted as a kind of "non-Euclidean metric" which makes the connection between minimizing (4.1) and computing an empirical Fréchet mean from the observations, see [28, 10] for further details.

The second interpretation comes from the statistical modeling of the template estimation problem. The energy (4.1) can be seen as the log likelihood of the observations and the deformations in the model presented in Eq. (3.9) given the parameters θ. Minimizing this energy is equivalent to maximizing the likelihood of both the observations and the random deformations.

Note that this optimization algorithm tries to find the best deformations which fit the template to the observations. In the discrete setting introduced in Subsection 2.2, we have shown that the diffeomorphic deformation is parameterized by both the initial momenta and the initial locations of the control points. This allows for optimizing with respect to both of these parameters. This optimization is detailed in [20] where the number of control points is also optimized using an L^1 penalty on the momenta.

4.2. **Stochastic approach**

The deterministic approaches aims at estimating the template through the estimation of the deformations first. Although these technics produce really impressive results on different databases, the lower bound (3.3) suggest to try some other approaches when the noise level increases. The statistical model (3.9) where the deformations are unobserved seems to be an appropriate answer. Whereas the deterministic approach computes the complete likelihood (observations together with random deformations), one may rather focus on the observation likelihood only. This means to compute the maximum of the the marginal of the previous quantity with respect to the deformation coefficients. This falls naturally into the case for which the Expectation-Maximization (EM) algorithm was developed.

Let us briefly recall the EM steps. This algorithm is an iterative algorithm which aims at maximizing

$$h(\theta) = \log\left(\int q(Y,\phi;\theta)d\phi\right). \tag{4.2}$$

Starting from an initialization of the parameter sequence θ_0, it creates a sequence of estimates $(\theta_k)_{k\in\mathbb{N}}$. First, it computes the expectation (E-step) of the complete log likelihood with respect to the posterior distribution of the unobserved data using the current parameters : $Q(\theta|\theta_k) = \mathbb{E}\left[\log q(Y,\phi;\theta)|Y;\theta_k\right]$. The second steps updates the parameter by maximizing (M-step) the function Q :

$$\theta_{k+1} = \operatorname*{argmax}_{\theta} Q(\theta|\theta_k).$$

A wide range of model falls into the so-called exponential family as the complete log-likelihood has the form $\log q(Y,\phi;\theta) = \langle S(Y,\phi), \psi(\theta)\rangle + \Phi(\theta)$ where S is a vector value function called the sufficient statistics and ψ and Φ are two measurable functions. This particular forms enables to simplifies the computations since the E-step only requires to compute the expectation of the sufficient statistics. Although quite complex, the deformable template model (3.9) belongs to this family. In this setting, the maximization step is explicit as a function of the sufficient statistics : $\theta_{k+1} = \hat{\theta}(s_k)$ where s_k is the conditional expectation of S at iteration k.

Unfortunately, the posterior distribution does not have a simple form so that the expectation cannot be calculated. Therefore, several methods have been proposed to reach an estimate of the parameters by approximating this expectation. One option is the deterministic algorithm presented above where the expectation is approximated by a single realization of the integrant taken at the mode of the posterior distribution. Another approach leads to the use of a stochastic version of the EM algorithm called Stochastic Approximation EM (SAEM) [19] coupled with Monte Carlo Markov Chain [5]. The E-step is replaced by a simulation of the non-observed random variables, namely the deformations, from an ergodic Markov chain transition kernel. A stochastic approximation is then performed on the sufficient statistics using these samples. The M-step is unchanged.

The almost sure convergence of this algorithm towards the MAP estimator has been proved in different situations, first in [5] for the Bayesian Mixed effect model presented in Subsection 3.2.2 and in [3] for the mixture

of this model detailed in Subsection 2.3. The requirements for convergence concern the Markov chain which has to be geometrically ergodic on any compact subset. Several choices may be done to satisfy this criterion [4, 2]. In [47], some particular representations of deformations and templates are introduced using a common finite element decomposition of the image domain. The resulting deformation and template fields have Markovian properties facilitating samplings and accelerating algorithm convergence.

When the data are observed sequentially, one can use the online EM [16]. Using a sequential algorithm is useful for several reasons: storing the observations throughout the process (which might be resource intensive in case of high dimensional data) is no longer required in an online setting, because each observation is processed only once. Then, when the E-step conditional expectation in intractable, as it is the case in the deformable template model, its approximation is much lighter than in a batch algorithm. Finally, in cases where the observed data evolves during the acquisition time, the sequential learning provides an evolution of the trend of the parameters throughout the iterations.

Under some mild assumptions, verified by the deformable template model, it is shown that the sequence of estimates $(\theta_k)_{k\in\mathbb{N}}$ of θ converges with probability one to the set of stationary points of the Kullback-Leibler divergence between the marginal distribution of the observation and the model distribution.

Compared to the SAEM algorithm detailed earlier, the difference lies in that the stochastic approximation involves the conditional expectation of the sufficient statistics only given the last available observation. This online EM can be also coupled with MCMC methods when the conditional expectation $\mathbb{E}[S(Y,\phi)|Y_{k+1};\hat{\theta}_k]$ is intractable.

We can notice that at iteration k of the online EM, a single chain targeting the posterior distribution of the deformation $q(\,\cdot\,|Y_{k+1};\hat{\theta}_k)$ is required, whereas n chains targeting the posterior distributions $q(\,\cdot\,|Y_1;\hat{\theta}_k)$, $\ldots, q(\,\cdot\,|Y_n;\hat{\theta}_k)$ are necessary in the SAEM algorithm, N denoting the total number of observations.

However, as already noticed in [3], the mixture model suffers from its sensitivity to the sampler which can lead to trapping states (states where the numerical value of the acceptation ration is too low so that the chain does not mix). In [17], it is suggested to slightly modify the original model (Section 2.3). The sampling scheme is adapted in the way that a Markov

Chain targeting the joint posterior distribution of the class and the deformations is run on an extended state space with a Metropolis within Gibbs algorithm. In this way, the class sampling takes advantage of all the extended hidden data and will allow to jump from one model to another, provided that consistent deformations are proposed. With this sampling scheme, the posterior simulation of the class random variable is performed more efficiently as the chain visits more different models.

5. **Applications**

In this section, we present some of the results on template estimations from the deformable template model.

5.1. **Estimated templates**

The first application is based on the USPS database. Although very simple images, this example enables to see the challenges of this issue. In particular, it enables to clearly see some templates which look relevant, it also allows for reconstructing synthetic samples from the estimated model to evaluate the geometric variability which has been captures. In addition, since a test database is available, it can be used as a classifier where each new image is classified in the class with maximum likelihood. The images are 16×16 grey level images. Fig. 5.1 presents the noisy training sample used to estimate the ten templates on the right hand side. For more detailed results on this example, see [1, 5, 3].

Another 3D medical training set (left rows of Fig. 5.2) has been tested. These binary images represent some murine dendrite spline excrescences which were generated on genetically modified mice to mimic the Parkinson's disease. The template is presented in Fig. 5.3 and some synthetic examples drawn with respect to the deformable template model using the estimated parameters are shown in the two right rows of Fig. 5.2. For more details on this experiment, see [4, 2].

The same databases where used to estimate multicomponent atlases when assuming that the population contained two classes. The results are presented on the right panel of Fig. 5.3 for the murine dendrite spine excrescences. These examples show the importance to be able to cluster the data in a given population. Indeed, the templates are very different from each other and therefore summarize much more precisely the population.

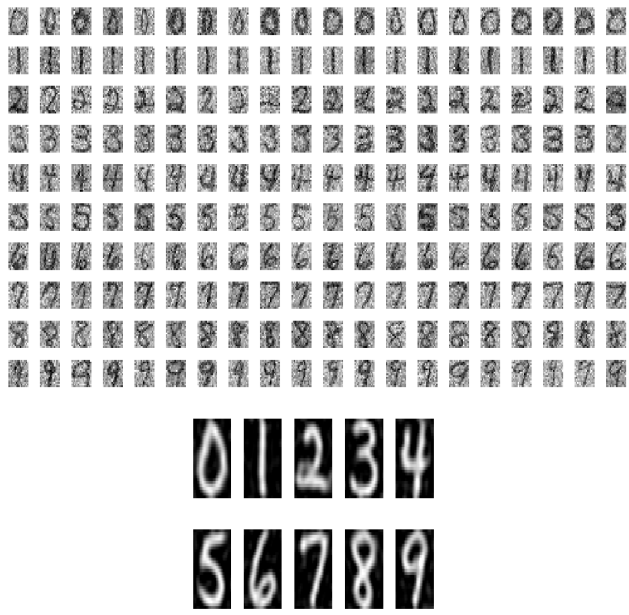


Figure 5.1. Top: exemplars of the USPS training database. Bottom: the ten templates estimated on this USPS base.

5.2. **Clustering and Classification**

5.2.1. **Classification via maximum a posteriori**

Thanks to the generative model presented in Section 3.2.2, one can classify new data calculating its likelihood to belong to different classes. After estimating the parameters of a deformable template model for each class, classification should be performed by computing the maximum posterior on class given the image. This classification has been performed on the USPS (c.f. Tables 5.1). The classification results show the importance of having enough but not too many images per components and to estimate the complete atlas namely the template as well as the geometrical variability (through the covariance matrix of the deformation distribution). When allowing for at most 15 components per class (some may be empty) and training on the whole available training sample of USPS (7291 images), the classification performances reaches an error of 3.5%. The estimation has

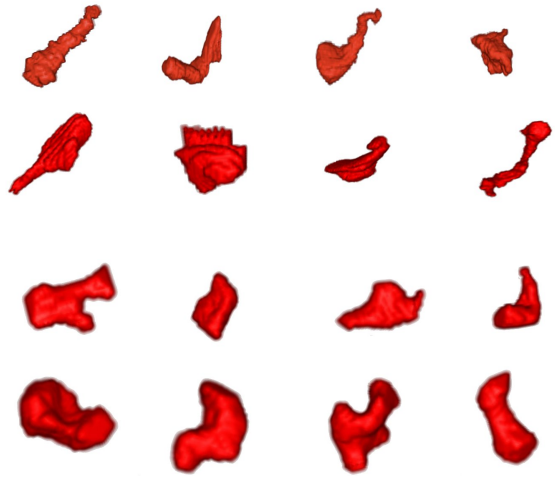


FIGURE 5.2. Top: Exemplars of the binary images of the segmented murine dendrite spine excrescences. Bottom: Synthetic samples generated from the deformable template model with the estimated parameters.

FIGURE 5.3. Left: Grey level template estimated from the 3D murine dendrite spine excrescences. Binary volume (second image) is created by thresholding the previous continuous image. (Third and forth images) Two spine excrescences templates estimated from the mixture deformable template model.

been done with the template estimated by the deterministic algorithm (FAM-EM); however, the results are similar with the other estimation algorithm (MCMC-SAEM) since, in the noise free case, the estimated parameters are very close to each other. In the noisy case, however, the stochastic algorithm outperforms the other estimation process which can clearly be seen on the parameter estimates ([5]).

A second test set is based on infrared images of combat aircrafts from the French Aerospace Lab ONERA. We refer to [38] for more details on the data and the classification performances.

Nb. of components	1	3	10
20 per class	7.36 (1.04)	7.78 (1.02)	13.71 (5.6)
40 per class	7.93 (1.06)	6.73 (0.78)	7.075 (2)
100 per class	9.4 (0.56)	5.48 (0.56)	5.43 (0.69)

Protocol	Photo. +Geom.	Sample mean +Geom.	Photo. + Geom. Prior
Mean error	7.36	19.80	18.22
Std of error	1.04	1.38	1.05

Table 5.1. Top: Error rates using the deterministic version of EM algorithm for different numbers of components (column) and different numbers of training images (rows) per class. Mean error rate and std. for 30 runs with randomly selected training sets. Bottom: Comparison of partial training protocols to full training protocol to highlight the importance of both photometrical and geometrical aspects (20 images per class, one cluster). "Photo" stands for estimation of photometric parameters. "Geom" stands for estimation of geometric parameters.

5.2.2. **Classifying registration**

The mixture model presented in Section 2.3 was used for the registration of mammograms in the context of computer-aided detection of breast cancer [34]. Usual automated detection techniques include comparisons of several mammograms either from left and right breasts of a same patient, from a same breast at different dates, or from a same breast before and after a contrast agent injection. In general, such techniques aim at localizing abnormal image differences which are signs of lesions. However, their application requires corrections of other differences due to normal factors. This is illustrated on images (a) to (c) of Figure 5.4 which show a pair of bilateral mammograms and their difference image. In image (c), observed differences result not only from the unilateral lesion (a circular opacity in the top left part of breast) but also from variations of breast size and

positioning (particularly visible near the bottom left part of the breast boundary).

In this context, image registration can be used as a mean to compensate for some normal differences. However, corrections made by registration should not concern abnormal differences. Indeed, these differences should be preserved for lesion detection to remain possible on registered images. But, usual registration techniques do not fulfill this specific requirement. Relying upon the assumption that intensity dependencies are spatially invariant, they do not distinguish between differences on normal and abnormal tissues, and correct both of them.

This shortcoming can be overcome using the mixture model in Section 2.3 with two classes ($K = 2$), one for normal tissues and another for lesions. As shown on images (e) and (f) of Figure 5.4, this model tends to enhance abnormal differences by reducing most exclusively the normal ones. Moreover, combining pixel classification and image registration, it also produces a lesion map (see image (d)) which can be directly used for detection.

5.3. **Spatiotemporal templates**

The spatiotemporal model (2.7) presented in Section 2.4 was successfully applied to dynamic contrast enhanced sequences of computed tomography images (DCE-CT sequences) [33]; some 2D-slices taken from a DCE-CT sequences are shown on the first row of Figure 5.5. This modality relies upon the injection of a contrast agent into the body. It is intended to observe during a short time period the diffusion of the agent through a body part, and to access to absorption properties of tissues. It is particularly used in cancerology for the diagnosis and the management of tumors, which are characterized by specific absorption properties.

The interpretation of DCE-CT signals is often ambiguous due to patient movements which produce geometrical variations interfering with contrast variations. Hence, so as to assist the radiologist in his analysis, it is particularly useful to develop computer tools which can separate both sources of variations, and compensate for the geometric ones. The spatiotemporal model (2.7) can be used to achieve this task. The temporal template that it enables to estimate is a version of the sequence which is attenuated in noise and compensated for movements. For illustration, an estimated temporal template is presented in the second row of the Figure 5.5.

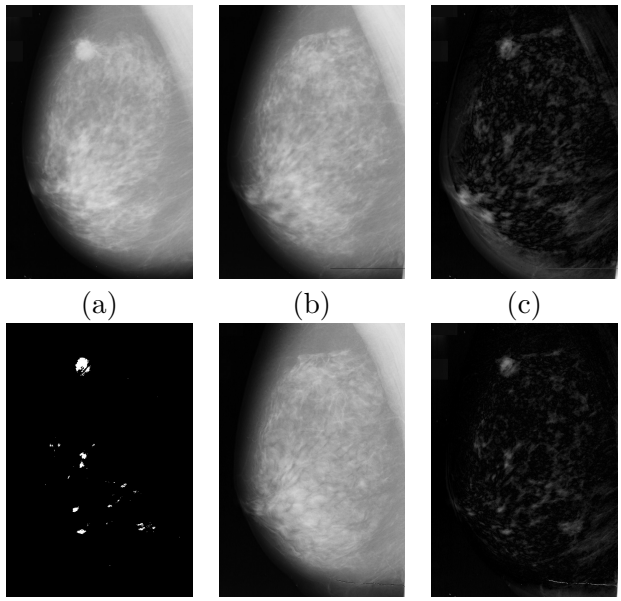


Figure 5.4. Images (a) and (b) form a pair of bilateral mammograms of a same woman. Image (c) is the difference between images (a) and (b) before registration. The image (e) is the deformation of image (b) obtained by registration of image (a) and (b) with a classifying registration technique. Image (f) is the difference between registered images. Image (d) is the class map estimated by a classifying registration technique.

6. **Conclusion**

This article summarizes the statistical aspects of the deformable template model in different mathematical frameworks. The weakness of this atlas estimation are of two types. First, the model itself can be optimized in particular concerning the deformation parameter distribution. The Gaussian prior, however very convenient, is not always realistic. Other distributions should be investigated. The other weakness concerns the numerical aspects. Indeed, due to the high dimension of the data and their increasing

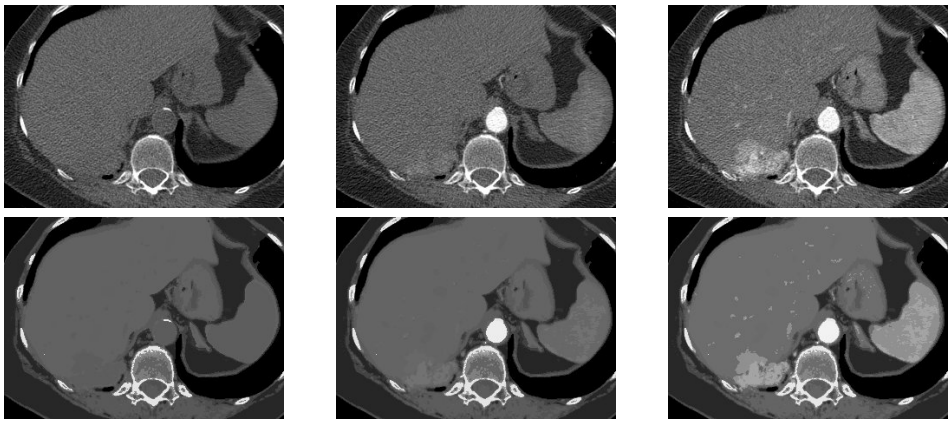


Figure 5.5. Abdominal sequence. Some images of original sequence (1st row) and the estimated template (2nd row).

number, the computational power of the algorithms has to be optimized particularly concerning the covariance matrix estimation. The estimation of this high dimensional full matrix has to be carefully considered in order to avoid memory issues.

However, this model features also major advantages. First, it enables to describe a population in terms of statistical descriptors such as mean and covariance. It also approximates the generation of these images or shapes which allows to better understand the different aspects -geometric and photometric- of the training set. Moreover, the generative form of the model enables to re-sample new elements, providing an augmented basis where subtle features may appear clearer. A second advantage of this model and its estimation algorithms is the theoretical properties it achieves. The consistence of the MAP estimators as well as the convergence of the estimation algorithms are crucial as they ensure the relevance of the results. Moreover, the theoretical bounds on deterministic procedures also provides information about the expected accuracy of estimates. The last major advantage of this model is that it can be generalized for many different problems. For example, including the segmentation issue in a probabilistic atlas estimation, dealing with multimodal databases or forcing the deformation to be diffeomorphic are directions of current researches which are all based on the described model here.

Other challenges remain open. Dealing with in depth the convergence issue that the deterministic algorithm has to face would enable to propose alternative options to the user depending on the data set he / she is considering. This aspect is of great importance as it would also make the link with the usual variational approaches and its validation. The sensitivity of outliers is also important and particularly in the online estimation process. These data should be detected along with the estimation and split up from the rest. This is partially done by the multi-templates model although selecting the number of components in advance remains difficult. The very high dimension of the variables is also a challenge. Reducing their dimension and more particularly adapting the parametric models should be part of the estimation. Finding variables of lower dimension may enable an easier interpretation of the estimates and also increase the classification power.

Many other works are related to the issue of template estimation using different models (e.g. [51, 43, 8, 54]) or algorithms ([7]). The results recalled here may be used to achieve theoretical properties of these approaches which, to the best of our knowledge, do not exists yet.

References

[1] S. Allassonnière, Y. Amit & A. Trouvé – "Toward a coherent statistical framework for dense deformable template estimation", *JRSS* **69** (2007), p. 3–29.

[2] S. Allassonnière & E. Kuhn – "Convergent stochastic expectation maximization algorithm with efficient sampling in high dimension. application to deformable template model estimation", *submitted*.

[3] ———, "Stochastic algorithm for bayesian mixture effect template estimation", *ESAIM Probab.Stat.* **14** (2010), p. 382–408.

[4] S. Allassonnière, E. Kuhn & A. Trouvé – "Bayesian consistent estimation in deformable models using stochastic algorithms: Applications to medical images", *Journal de la Société Française de Statistique* **151(1)** (2010), p. 1–16.

[5] ———, "Bayesian deformable models building via stochastic approximation algorithm: A convergence study", *Bernoulli J.* **16** (2010), no. 3, p. 641–678.

[6] S. Allassonnière, A. Trouvé & L. Younes – "Geodesic shotting and diffeomorphic matching via textured meshes", in *Proc. of the Energy Minimization Methods for Computer Vision and Pattern Recognition (EMMCVPR 05)* (A. Y. Anand Rangarajan, Baba Vemuri, éd.), November 9-11 2005, p. 365–381.

[7] J. Ashburner – "A fast diffeomorphic image registration algorithm", *NeuroImage* **38** (2007), p. 95–113.

[8] J.-Y. Audibert, G. Charpiat, O. Faugeras & R. Keriven – "Imaes statistics based on diffeomorphc matching", Tech. report, ENPC, Fev 2005.

[9] M. F. Beg, M. I. Miller, A. Trouvé & L. Younes – "Computing large deformation metric mappings via geodesic flows of diffeomorphisms", *Int J. Comp. Vis.* **61** (2005), no. 2, p. 139–157.

[10] J. Bigot & B. Charlier – "On the consistency of Fréchet means in deformable models for curve and image analysis", *Electron. J. Stat.* **5** (2011), p. 1054–1089.

[11] J. Bigot, C. Christophe & S. Gadat – "Random action of compact lie groups and minimax estimation of a mean pattern", *Information Theory, IEEE Transactions on* **58** (2012), no. 6, p. 3509 –3520.

[12] J. Bigot & S. Gadat – "A deconvolution approach to estimation of a common shape in a shifted curves model", *Ann. Statist.* **38** (2010), no. 4, p. 2422–2464.

[13] J. Bigot, F. Gamboa & M. Vimond – "Estimation of translation, rotation, and scaling between noisy images using the Fourier-Mellin transform", *SIAM J. Imaging Sci.* **2** (2009), no. 2, p. 614–645.

[14] J. Bigot, J.-M. Loubès & M. Vimond – "Semiparametric estimation of shifts on compact lie groups for image registration", *Probab. Theory Related Fields* (2010), p. 1–49.

[15] J. Bigot, S. Gadat & J.-M. Loubes – "Statistical M-estimation and consistency in large deformable models for image warping", *J. Math. Imaging Vision* **34** (2009), no. 3, p. 270–290.

[16] O. Cappé & E. Moulines – "Online EM Algorithm for Latent Data Models", *J. R. Statist. Soc.B* **71** (2007), p. 593–613.

[17] B. P. Carlin & S. Chib – "Bayesian model choice via Markov chain Monte Carlo", *J. R. Statist. Soc.B* **57** (1995), p. 473–484.

[18] G. E. Christensen, R. D. Rabbitt & M. I. Miller – "Deformable templates using large deformation kinematics", *IEEE trans. Image Proc.* **5** (1996), no. 10, p. 1435–1447.

[19] B. Delyon, M. Lavielle & E. Moulines – "Convergence of a stochastic approximation version of the EM algorithm", *Ann. Statist.* **27** (1999), no. 1, p. 94–128.

[20] S. Durrleman, S. Allassonnière & S. Joshi – "Sparse adaptive parameterization of variability in image ensembles", *International Journal of Computer Vision* **101(1)** (2013), p. 161–183.

[21] S. Durrleman, X. Pennec, A. Trouvé & N. Ayache – "Statistical models on sets of curves and surfaces based on currents", *MedIA* **13** (2009), no. 5, p. 793–808.

[22] S. Durrleman, X. Pennec, A. Trouvé, G. Guerig & N. Ayache – "Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets", *proc. of the* MICCAI conf. (2009).

[23] J. Fishbough, S. Durrleman & G. Guerig – "Estimation of smooth growth trajectories with controlled acceleration from time series shape data", *proc. of the* MICCAI conf. (2011).

[24] M. Fréchet – "Les éléments aléatoires de nature quelconque dans un espace distancié", *Ann. Inst. H.Poincaré, Sect. B, Prob. et Stat.* **10** (1948), p. 235–310.

[25] F. Gamboa, J.-M. Loubes & E. Maza – "Semi-parametric estimation of shifts", *Electron. J. Stat.* **1** (2007), p. 616–640.

[26] C. A. Glasbey & K. V. Mardia – "A penalized likelihood approach to image warping", *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** (2001), no. 3, p. 465–514.

[27] J. Glaunès, M. Vaillant & M. I. Miller – "Landmark matching via large deformation diffeomorphisms on the sphere", *Journal of Mathematical Imaging and Vision, MIA 2002 special* **20** (2004), p. 179–200.

[28] J. Glaunès & S. Joshi – "Template estimation form unlabeled point set data and surfaces for computational anatomy", in *Proc.*

of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA) (X. Pennec & S. Joshi, éds.), 2006, p. 29–39.

[29] J. Glaunès, A. Qiu, M. Miller & L. Younes – "Large deformation diffeomorphic metric curve mapping", *International Journal of Computer Vision* **80** (2008), no. 3, p. 317–336.

[30] J. Glaunès, A. Trouvé & L. Younes – "Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and submanifolds matching", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2** (2004), p. 712–718.

[31] U. Grenander – *General pattern theory*, Oxford Science Publications, 1993.

[32] U. Grenander, M. I. Miller & A. Srivastana – "Hilbert-Schmidt lower bounds for estimators on matrix lie groups for atr", *IEEE Trans. Pattern Anal. Mach. Intell.* **20** (1998), p. 790–802.

[33] M. Hachama, A. Desolneux, C. Cuénod & F. J. Richard – "A classifying registration technique for the estimation of enhancement curves of DCE-CT scan sequences", *Medical Image Analysis* **14** (2010), no. 2, p. 185–194.

[34] M. Hachama, A. Desolneux & F. J. Richard – "A bayesian technique for image classifying registration", *IEEE Transactions on Image Processing* **21** (2012), no. 9, p. 4080–4091.

[35] R. Holm, D, T. Ratnanather, J, A. Trouvé & L. Younes – "Soliton dynamics in computational anatomy", *Neuroimage* **23** (2004), p. S170–S178.

[36] S. Joshi & M. Miller – "Landmark matching via large deformation diffeomorphisms", *IEEE transactions in image processing* **9** (2000), no. 8, p. 1357–1370.

[37] S. Joshi, B. Davis, M. Jomier & G. Gerig – "Unbiased diffeomorphic atlas construction for computational anatomy", *Neuroimage* **23** (2004), p. S151–S160.

[38] S. Lefebvre, S. Allassonnière, J. Jakubowicz, T. Lasne & E. Moulines – "Aircraft classification with a low resolution

infrared sensor", *Machine Vision and Application Journal* **24(1)** (2012), p. 175–186.

[39] M. Lorenzi, N. Ayache, G. B. Frisoni & X. Pennec – "Mapping the effects of $A\beta_{1-42}$ levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields", in *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, LNCS, Springer, 2011, p. 663–670.

[40] M. Lorenzi, N. Ayache & X. Pennec – "Schilds ladder for the parallel transport of deformations in time series of images", in *Proceedings of Information Processing in Medical Imaging (IPMI'11)* (G. Szekely & H. Hahn, éds.), LNCS, vol. 6801, 2011, Honorable Mention (runner-up) for the Erbsmann Award, p. 463–474.

[41] F. Maire, S. efebvre, E. Moulines & R. Douc – "An online learning algorithm for mixture models of deformable templates", in *Proc. of the :2012 IEEE Workshop on Machine Learning for Signal Processing*, 2012.

[42] A. Makadia & K. Daniilidis – "Rotation recovery from spherical images without correspondences", *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28** (2006), no. 7, p. 1170–1175.

[43] S. Marsland & C. Twining – "Constructing diffeomorphic representations for the groupewise analysis of non-rigid registrations of medical images", *IEEE Transactions on Medical Imaging* **23** (2004).

[44] I. Miller, M & L. Younes – "Group action, diffeomorphism and matching: a general framework", *Int. J. Comp. Vis* **41** (2001), p. 61–84, (*Originally published in electronic form in: Proceeding of SCTV 99, http://www.cis.ohio-state.edu/ szhu/SCTV99.html*).

[45] W. Park, D. R. Madden, D. N. Rockmore & G. S. Chirikjian – "Deblurring of class-averaged images in single-particle electron microscopy", *Inverse Problems* **26** (2010), no. 3, p. 035002, 29.

[46] A. Qiu, M. Albert, L. Younes & M. I. Miller – "Time sequence diffeomorphic metric mapping and parallel transport track time-dependent shape changes", *NeuroImage* **45** (2009), p. 51–60.

[47] F. J. Richard, A. Samson & C. A. Cuenod – "A SAEM algorithm for the estimation of template and deformation parameters in medical image sequences", *Statistics and Computing* **19** (2009), p. 465–478.

[48] M. Sabuncu, S. K. Balci & P. Golland – "Discovering modes of an image population through mixture modeling", *MICCAI* **LNCS** (2008), no. 5242, p. 381–389.

[49] D. W. Thompson – *On growth and form*, Cambridge : University Press ; New York : Macmillan, 1915.

[50] A. Trouvé – "Infinite dimensional group action and pattern recognition", Tech. report, DMI, Ecole Normale Supérieure, 1995.

[51] C. Twinings, S. Marsland & C. Taylor – "Measuring geodesic distances on the space of bounded diffeomorphisms", in *British Macine Vision Conference*, 2002.

[52] A. W. van der Vaart – *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, Cambridge, 1998.

[53] M. Vaillant & J. Glaunès – "Surface matching via currents", in *Information Processing in Medical Imaging* (G. Christensen & M. Sonka, éds.), Lecture Notes in Computer Science, vol. 3565, Springer Berlin / Heidelberg, 2005, 10.1007/11505730_32, p. 1–5.

[54] T. Vercauteren, X. Pennec, A. Perchant & N. Ayache – "Diffeomorphic demons: Efficient non-parametric image registration.", *Neuroimage* **45** (2009), p. 61–72.

[55] F. c.-X. Vialard & A. Trouvé – "A second-order model for time-dependent data interpolation: Splines on shape spaces", *STIA Workshop Miccai* (2010).

[56] M. Vimond – "Efficient estimation for a subclass of shape invariant models", *Ann. Statist.* **38** (2010), no. 3, p. 1885–1912.

Statistics on Deformable templates

Stéphanie Allassonnière
CMAP
Ecole Polytechnique
Route de Saclay
91128 Palaiseau
FRANCE
stephanie.allassonniere@polytechnique.edu

Jérémie Bigot
Institut de Mathématiques de Toulouse,
CNRS UMR 5219
Université de Toulouse
118 route de Narbonne
31062 Toulouse Cedex 9
FRANCE
jeremie.bigot@math.univ-toulouse.fr

Joan Alexis Glaunès
MAP5
Université Paris Descartes, Sorbonne
Paris Cité
45 rue des Saints-Pères
75270 Paris Cedex 06
FRANCE
alexis.glaunes@mi.parisdescartes.fr

Florian Maire
ONERA - The French Aerospace Lab
F-91761 Palaiseau
FRANCE
florian.maire@onera.fr

Frédéric J.P. Richard
LATP CNRS UMR 7353
Aix Marseille Université
Centre de mathématiques et
d'informatique
39 rue Frédéric Joliot
13453 Marseille Cedex
FRANCE
frederic.richard@univ-amu.fr