Wavelet-Based Multiscale Initial Flow For Improved Atlas Estimation in the Large Diffeomorphic Deformation Model Framework

Fleur Gaudfernau, Eléonore Blondiaux, Stéphanie Allassonnière, Erwan Le Pennec

Wavelet-Based Multiscale Initial Flow For Improved Atlas Estimation in the Large Diffeomorphic Deformation Model Framework

Fleur Gaudfernau[1*], Eléonore Blondiaux[3], Stéphanie Allassonnière[1] and Erwan Le Pennec[2,4]

[1]CRC, Université de Paris, INRIA EPI HeKa, INSERM UMR 1138, Sorbonne Université.
[2]CMAP, Ecole polytechnique, Institut Polytechnique de Paris.
[4]XPOP, Inria Saclay.
[3]Service de Radiologie, Hôpital Armand-Trousseau, APHP.

*Corresponding author(s). E-mail(s): fleur.gaudfernau@etu.u-paris.fr;

Abstract

Modelling the mean and variability in a population of images, a task referred to as atlas estimation, remains very challenging, especially in a clinical setting where deformations between images can occur at multiple scales. In this paper, we introduce a coarse-to-fine strategy for atlas estimation in the Large Deformation Diffeomorphic Metric Mapping framework, based on a finite parametrization of the subjects' velocity field. Using the Haar Wavelet Transform, a multiscale representation of the initial velocity fields is computed in order to optimize the template-to-subject deformations in a coarse-to-fine fashion. This reparametrization preserves the reproducing kernel Hilbert space structure of the velocity fields, enabling the algorithm to perform efficiently gradient descent. Numerical experiments on three different datasets, including a dataset of abnormal fetal brain images, show that compared to the original algorithm, the coarse-to-fine strategy reaches higher performance and yields template images that preserve important details while avoiding unrealistic features.

Keywords: Deformable template model, atlas estimation, diffeomorphic deformations, Haar Wavelet, coarse-to-fine algorithm

1 Introduction

Although the quantitative analysis of anatomical images is an old problem (Thompson, 1992), to this day it is still a challenging one, especially given that datasets of clinical images are often small in number and large in dimensionality. Estimating the transformation that warps a shape onto another provides an efficient way of quantifying shape differences, which is the cornerstone idea of Computational Anatomy (Grenander and Miller, 1998). When one works with collections of images that are instances of the same anatomical object, the question arises as to how faithfully model the mean and variability over these structures. Atlas estimation is a method to achieve such modeling: an estimate of the average shape is given in the form of a template image, which represents the invariants across the population, i.e. shared anatomical features, and the variability is given by deformations from the template space to each subject's space, which express how these common features vary within the population (Grenander and Miller, 1998).

Atlas estimation has many applications in the field of medical image analysis. The template image can be used as reference to describe typical anatomical structures or serve as a tool to automatically segment new subjects. Variations around the template may be used to characterize pathological deviations from normality (Gaudfernau et al, 2021) or to isolate subgroups in the population (Debavelaere et al, 2020). Atlases can also be defined in a spatio-temporal fashion to characterize normal or pathological changes, such as brain growth across gestation (Gholipour et al, 2017) or hippocampus reduction in Alzheimer's disease (Debavelaere et al, 2020).

As in registration problems, the choice of the deformation function describing the template-to-subject transformations is of primary importance (Oliveira and Tavares, 2014). To account for the intra- and inter-subject anatomical variability in clinical images, nonlinear deformations are mandatory. Among these, diffeomorphisms are a good choice of transformation as they are high-dimensional, smooth and invertible functions with smooth inverse that preserve the topology of anatomical images.

The Large Deformation Diffeomorphic Metric Mapping (LDDMM) setting (Trouvé, 1998; Miller et al, 2002; Christensen et al, 1996) is a mathematical framework in which objects are deformed through diffeomorphic transformations of the whole ambient space. The group of possible transformations forms a Riemannian manifold of infinite dimension and parametrize a flexible representation of deformations. Such models have a very high number of degrees of freedom and parameters, hence the need to narrow the search space to a small subgroup of deformations. This is typically done by optimizing a cost function that comprises a distance between the deformed template and each subject, plus an energy term acting as spatial regularizer (Durrleman et al, 2012). The latter constrains diffeomorphisms to be geodesics on the Riemanian Manifold, i.e. the shortest paths between the identity map and the diffeomorphism of interest according to a regularizing metric.

In the LDDMM framework, diffeomorphisms are constructed by integrating time-dependent velocity fields. Conveniently, the resulting flow of diffeomorphisms is fully determined by the initial state of the system, which the model seeks to optimize. The question of how to parametrize this deformation field has been subject to significant research. Notably, Durrleman et al (2012) proposed a finite and sparse parametrization of velocity fields based on the convolution of momentum vectors attached to control points. Velocity fields are usually regularized by a Gaussian kernel, which defines a Reproducing Kernel Hilbert Space (RKHS) and induces a right-invariant metric on the space of diffeomorphisms (Durrleman et al, 2012).

The choice of this regularizer is critical as it restricts the range of transformations defined by the model (Sommer et al, 2012). Specifically, it constrains the deformations occurring on the images to a single scale. A large kernel width is likely to produce smooth but inaccurate deformations while a fine kernel will generate more accurate but noisy deformations. As clinical images often present high variability at several scales, one might be tempted to increase the number of parameters in the model by using many control points and a small kernel. However, fine kernels make large displacements more expensive than small ones, and such over-parametrization will likely trap the optimization procedure in a local minimum, achieving a reasonable numerical solution that is qualitatively bad, especially in terms of template realism. To overcome such problem, hierarchical algorithms have been widely used in the field of image registration (Oliveira and Tavares, 2014; Modersitzki, 2008): they first solve the registration problem at coarse scales, and then refine the transformation at finer and finer scales by transferring the solution from one level to the next. These strategies can avoid more efficiently trapping the algorithm in local minima related to irregular transformations. Except for Tan and Qiu (2016), they have yet to be applied in the LDDMM framework.

In this paper, we propose a coarse-to-fine atlas estimation procedure based on a wavelet representation of the initial velocity fields. This strategy has the advantage of rendering the algorithm less dependent on the initialization while favoring more multiscale, hence natural deformations. We consider the LDDMM setting and we rely on the finite parametrization of the velocity fields as a linear combination of RKHS basis elements presented in Durrleman et al (2012). Importantly, our strategy enables us to preserve this structural assumption and the efficient numerical scheme that

follows. We will show that our algorithm promotes more natural template images as well as higher stability regarding the initialization.

This paper is organized as follows. We first explore works related to coarse-to-fine registration and atlas estimation in Section 2, then we recall the LDDMM setting in the case of the finite parametrization of the velocity fields in Section 3 and introduce our coarse-to-fine atlas estimation method in Section 4; then, we conduct experiments in Section 5 and discuss our results in Section 6.

2 Related work

Even though multiscale image registration has been studied repeatedly in the literature, it has rarely been extended to population analysis. As registration is a special case of atlas estimation with a fixed template image, in the following we will review both registration and atlas estimation methods that have a multiscale property.
Multiscale strategies differ in the way they handle the different scales of the optimization: some methods favor a coarse-to-fine fashion while others handle coarse and fine scales simultaneously. We will first review coarse-to-fine strategies with a focus on those relying on a wavelet-based decomposition. Then, we will survey the multiscale strategies that have been introduced in the LDDMM framework.

Wavelet-based multiscale registration. Coarse-to-fine or hierarchical optimization strategies seek to solve the registration problem at progressively increasing resolutions, with the objective of decreasing the computational cost and finding a more accurate solution (Modersitzki, 2008). The search area is first restricted to coarse functions and the results are progressively refined, with the parameters estimated at the previous coarser level propagated to the next finer level. Such strategies can be coarse-to-fine with regard to the deformation field (Debroux et al, 2021; Downie et al, 1996; Wu et al, 2000; Gefen et al, 2004; Sun et al, 2014), and/or with regard to the registered images (Debroux et al, 2021; Paquin et al, 2006, 2007, 2008; Athavale et al, 2015; Olivo et al, 1995; Tadmor and Nezzar, 2004). We will not dwell on the latter case, which is out of the scope of this paper.

The Wavelet Transform provides a multiresolution decomposition of signals into a basis of functions with specific frequency, location and orientation (Mallat, 1989). Applying this hierarchical decomposition to deformations provides a convenient basis for hierarchical strategies. Amit (1994) first showed that modelling deformations using a wavelet basis provides better spatial regularization compared to using a Fourier basis. Wavelet-based decomposition algorithms were extended to a variety of wavelet types, such as Haar (Downie et al, 1996), Cai Wang (Wu et al, 2000; Sun et al, 2014) and (BV, L^2) (Debroux et al, 2021), and to a variety of deformations, such as motion vectors (Wu et al, 2000), B-splines (Sun et al, 2014) and elastic deformations (Debroux et al, 2021). Of note, the Wavelet Transform has also been used in deformation analysis with various aims such as data visualization and compression, again encompassing various types of wavelet bases such as the Cauchy-Navier wavelet (Mk et al, 2003), the non-linear Morphlet wavelet (Kaplan and Donoho, 2021) and the spline wavelet (Lam et al, 2017).

Multiscale registration in the LDDMM framework In the LDDMM framework, the choice of the spatial regularizer restricts the range of possible deformations to those occurring at a single scale, which often proves unrealistic (Risser et al, 2011). Thus, a variety of papers have focused on increasing the flexibility of the deformation model. These strategies can be broadly classified into two categories: the first one simultaneously estimates coexisting flows of different scales (Risser et al, 2011; Bruveris et al, 2011; Sommer et al, 2012; Gris et al, 2016; Tan and Qiu, 2016), and the second one composes multiple scale flows which are estimated sequentially (Modin et al, 2018; Miller et al, 2020).

Risser et al (2011); Bruveris et al (2011) first introduced a multi-kernel extension of the LDDMM framework in which the deformation flow is defined by a weighted sum of Gaussian kernels whose widths are specified by the user. Weights are tuned in a semi-automatic manner during a pre-registration step. A spatially-varying version of this multi-kernel framework, the kernel bundle, used sparsity priors to allow the weights of the kernel mixture to vary across spatial locations (Sommer et al, 2012). In this framework, the RKHS structure of the velocity fields is lost, and a new definition of the norm was introduced to ensure efficient computation of the flow. Even though this algorithm proved efficient on the registration of landmark points,

the increase in computational cost restricts its application to registration problems involving few parameters. The kernel bundle framework was also combined with a coarse-to-fine strategy in an algorithm for cortical surface registration (Tan and Qiu, 2016): the resolution of the surfaces is progressively increased along with that of the deformation field. This strategy has yet to be applied to the registration of images. Multi-kernel approaches were further combined with deep learning optimization in order to learn a local regularizer from the data (Niethammer et al, 2019; Shen et al, 2019). However, these methods increase significantly the complexity of the mathematical model, and several optimization procedures are required to tune the networks parameters, the kernel pre-weights and the deformation parameters.

As in the kernel bundle framework, the tuning of additional parameters is heavy on the computational cost, which dampens the application of such algorithms to high-dimensional images such as volumetric Magnetic Resonance Images (MRIs). Of note, a related approach, based on modular deformations, enables the user to impose spatially-varying constraints on the deformation field (Gris et al, 2016). Large diffeomorphic deformations are built by superimposing deformations modules, which encode local geometrical transformations, making it possible to construct diffeomorphisms from multiple scale flows. As in the kernel bundle framework, the space of vector fields is equipped with an adapted norm. The need for prior knowledge about the deformation modules limits the practical application of the algorithm.

A second and less explored axis of research constructed a hierarchical representation of deformations, based on non-coexisting vector flows of increasing resolution, which are estimated independently and then composed. In a theoretical paper, Modin et al (2018) extended (BV, L^2) wavelets to express diffeomorphisms as a composition of deformations of increasingly fine scales, which can be seen as a series of LDDMM steps. Despite the potential of this approach, it was not experimented numerically. A similar approach (Miller et al, 2020) constructed diffeomorphisms by composing a series of multiscale vector fields, which enable to progressively refine the deformation. Contrary to multi-kernel approaches, such strategies perform optimization in successive RKHS of decreasing smoothness, in the spirit of coarse-to-fine strategies.

As we shall see in the following, our coarse-to-fine approach is more closely related to the one that composes multiple scale flows, in the sense that we perform optimization sequentially in sub-spaces of increasing resolution. However, our algorithm differs from the previous ones mainly by the fact that almost no complexity is added to our mathematical model: the velocity fields are still defined, at core, by a single-scale RKHS, which simplifies the implementation of our algorithm while preserving the efficient optimization scheme of Durrleman et al (2012).

3 Model of diffeomorphic deformations

3.1 Large Deformation Diffeomorphic Metric Mapping

In the following, we consider a set of N images $(I_i)_{1\leq i\leq N}$ of dimension d. We assume that each image I_i is a smooth deformation of a template I_0 plus an additive random white noise ϵ_i:

$$I_i = I_0 \circ \Phi_i^{-1} + \epsilon_i, \ \forall i \in [1, n] \tag{1}$$

where Φ_i is the i^{th} template-to-subject deformation, and $I_0 \circ \Phi_i^{-1}$ denotes the action of the diffeomorphic deformation on the template. In the atlas estimation problem, one seeks to estimate the template intensities I_0 and the N template-to-subject deformations. Note that registration is a specific case of atlas estimation where the template image is fixed and $N = 1$.

We choose to work in the LDDMM setting (Trouvé, 1998; Miller et al, 2002), in which objects are deformed via deformations of the whole ambient space. This framework generalizes the linearized deformation setting in order to define diffeomorphic deformations that are invertible and smooth. Diffeomorphisms are constructed by integrating linearized deformations over time, which are considered as infinitesimal steps, according to the differential flow equation:

$$\begin{cases} \dfrac{dx(t)}{dt} = v_t(x(t)) \\ x(0) = x_0 \,. \end{cases} \tag{2}$$

This equation describes the motion of a particle x along the curve $x(t)$ in the domain of interest D. The vector field $(v(t))_t$ plays the role of an instantaneous velocity field. This model builds a flow of diffeomorphisms $\Phi_t : x_0 \longrightarrow x(t)\ \forall t \in [0,1]$. The diffeomorphism of interest Φ_1 is defined as the end point of the path $x(t)$, i.e.:

$$\forall x_0 \in D, \quad \Phi_1(x_0) = x(1)\,.$$

Note that $\forall t\ \Phi_t$ is indeed a diffeomorphism provided that the velocity field is regular enough.

To estimate the template intensities I_0 and the N template-to-subject deformations, we seek to minimize a cost function E composed of a trade-off between a data fidelity term quantifying how well the deformations of the template fit the observed images and a regularity term penalizing the kinetic energy of the deformation:

$$E = \sum_{i=1}^{N} \left(\frac{\|I_i - I_0 \circ \Phi_{1,i}^{-1}\|^2}{\sigma^2} + \int_{t=0}^{1} \|v_{t,i}\|_V^2 \right), \tag{3}$$

where σ controls the trade-off between the two terms. The first term is the sum of squared differences between the deformed template and each image. The second term imposes that the solutions are geodesic for the $\|.\|_V$ norm. Finally, the regularity in the RKHS is large enough for the diffeomorphisms to be well defined.

3.2 Discrete parametrization of diffeomorphisms

While the group of diffeomorphisms is of infinite dimension in theory, transformations can be restricted to a subgroup of deformations through a discrete parametrization of the subjects' velocity field (Durrleman et al, 2012). We restrict ourselves to vector fields that belong to a RKHS V, namely the set of square integrable functions convolved with a regularizing kernel. More precisely, we assume that any $v_{t,i}$ can be decomposed as a finite linear combination of the RKHS basis vector fields using a set of momentum vectors $(\alpha_{k,i})_k$ attached to k_g control points $(c_{k,i})_k$:

$$v_{t,i}(x) = \sum_{k=1}^{k_g} K_g(x, c_{k,i}(t))\alpha_{k,i}(t)\,, \tag{4}$$

where K_g is usually a Gaussian kernel whose width is denoted σ_g.

We now assume that the initial velocity field $v_{0,i}$ writes as the interpolation of momentum vector $\alpha_{k,i}(0)$ located at control points $c_{k,i}(0)$:

$$v_{0,i}(x) = \sum_{k=1}^{k_g} K_g(x, c_{k,i}(0))\alpha_{k,i}(0).$$

In Miller et al (2006), it is proved that the vector fields that are geodesics with respect to the norm $\int_0^1 \|v_{t,i}\|_V^2 dt$ keep the same structure along time and write according to Equation (4).

Furthermore, the evolution of the control point positions $(c_{k,i}(t))_k$ and momentum vectors $(\alpha_{k,i}(t))_k$ is described by the Hamiltonian system equations (Miller et al, 2006). Finally, one verifies that the kinetic energy along geodesic paths is preserved over time, i.e. $\forall t \in [0,1], \|v_{t,i}\|_V = \|v_{0,i}\|_V$. This implies that when $v_{t,i}$ is a geodesic, it is fully parametrized by the initial velocity field $v_{0,i}$. Hence, estimation of the diffeomorphism $\Phi_{1,i}$ boils down to a geodesic shooting problem. The system is deterministic, and we only need to optimize the initial conditions $\alpha_{0,i} = (\alpha_{k,i}(0))_k$ and $c_{0,i} = (c_{k,i}(0))_k$ for each subject i along with the template image I_0.

3.3 Optimization

In this work, the position of the control points $(c_{0,i})_i$ is fixed. With the parametrization introduced in the previous section, the cost function E of Equation (3) is a function of I_0 and $(\alpha_{0,i})_i$:

$$E(I_0, (\alpha_{0,i})_{1\leq i\leq N}) = \sum_{i=1}^{N} \left(\frac{\|I_i - I_0 \circ \Phi_{1,i}^{-1}\|^2}{2\sigma^2} + \|v_{0,i}\|_V^2 \right)$$

Even if we are not in a convex setting, optimization is performed through gradient descent. Thus, we need to compute the gradient with respect to all the parameters. This is not a trivial task, but an efficient numerical scheme has been proposed in Durrleman et al (2012). It relies heavily on the fact that the vector fields that are solutions of the problem always remain finite sums of kernels of the RKHS and that the norm of the vector fields is the one of the RKHS. This strong structural assumption allows resolving efficiently

the Hamiltonian equations of the flow, deform the template image with the flow and thus compute the gradient with respect to the template. Finally, using a backward integration along time, the gradient with respect to the momentum vectors can be computed.

This algorithm is publicly available as part of the open-source software Deformetrica[1].

3.4 Initialization, number of points and scale

The previous optimization problem is not convex, hence the gradient descent algorithm converges towards a solution that depends on the initialization. Furthermore, in Durrleman et al (2012), it is observed that the algorithm may not converge to a good solution when the number of control points and thus of parameters is too large. It should also be stressed that there is a dependency between the scale of the kernel and the number of parameters, as a constant vector field has to be well approximated by the finite sum: this imposes that the scale σ_g should be related to the distance between the points. Few control points imply a large σ_g, and therefore very regular vector fields, while many control points imply for a smaller σ_g and therefore more irregular vector fields.

In the original approach of Durrleman et al (2012), both the number and position of the control points are numerically optimized by applying a L1 sparsity prior to the momenta in order to inactivate points in zones with low variability and concentrate the remaining points in challenging areas. Unfortunately, the scale of the kernel remains the same and in practice, the vector fields can only be set to zero in empty areas. Further, the number and position of the control points can be optimized only if one uses a relatively low number of points, making the algorithm unable to cope with information-rich images.

One idea would be to change locally the scale of the kernel so that we can estimate non evenly smooth vector fields. Even if such parametrization can be written, one looses the RKHS structure and thereby the ease of computation. Note that Sommer et al (2012) has proposed a numerical scheme to cope with this setting where there is only a finite number of possible scales and for a very specific, and not natural, definition of the norm of the vector fields. Alternatively, in line with (Miller et al, 2020), we could solve the problem in several RKHS of increasingly fine kernel widths, but then again this would render the optimization scheme of (Durrleman et al, 2012) inapplicable.

In this paper, we wish to address these two related issues: the dependency of the algorithm on the initialization, which restrains the number of parameters that can be properly optimized, and the impossibility to estimate vector fields that have locally varying regularity. In the next section, we will describe a reparametrization of the vector fields which enables us to impose smoothness constraints on the deformations and progressively relax them in a coarse-to-fine fashion. In this way, the algorithm can cope with non evenly smooth transformations while using a small σ_g and remaining in the original RKHS setting.

4 Multiscale atlas estimation

Here, we introduce a new method for coarse-to-fine atlas estimation in which we use a multiscale representation of the initial velocity fields based on the Haar Wavelet Transform.

4.1 Reparametrization of the initial velocity field based on the Haar Wavelet Transform

Since our algorithm relies on the multiscale structure of the Wavelet Transform, we introduce the definition and properties of the continuous Haar Wavelet representation. We recall how this construction can be extended to a representation of discrete signals defined on a grid and we demonstrate how this can be used to obtain a Haar representation of the initial velocity fields.

4.1.1 Continuous Haar Wavelet Transform

Here, we describe the decomposition of a real signal f defined on the d dimensional space $\mathbb{R}^d$ into a Haar Wavelet basis (Mallat, 1989, 2008). The wavelet representation decomposes f into a linear combination of basis functions which have different resolution, location and orientation. This representation relies on a collection of embedded spaces V_s that contain functions said of scale s. In the case of the Haar Wavelet, V_s is the space of piecewise constant functions on a regular grid of size 2^s. Any function f can

[1] https://www.deformetrica.org/

be approximated in this space by computing a local average: the mean value in each subsquare of the grid. If we define the piecewise constant function

$$\psi^L(x) = \begin{cases} 1 & \text{for } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases},$$

and the d dimensional scaling function ϕ by

$$\phi(x) = \prod_{i=1}^{d} \psi^L(x_i)$$

with $x \in \mathbb{R}^d$, approximating f at scale s amounts to projecting f onto the space spanned by the orthonormal family

$$\left\{\phi_{s,k}(x) = 2^{-sd/2}\phi(2^{-s}x - k)\right\}_{k\in\mathbb{Z}}$$

where $\phi_{s,k}$ is the scaling function rescaled by 2^s and then translated by $k2^s$. The factor $2^{-sd/2}$ is a normalization factor that ensures that the $\phi_{s,k}$ have unit energy.

When transitioning from the approximation at scale s to the approximation at the coarser scale $s+1$, some details of f are lost. These details belong to the orthogonal complement W_{s+1} of the space V_{s+1} in V_s. A basis of this space can be obtained by defining the piecewise constant function

$$\psi^H(x) = \begin{cases} 1 & \text{for } 0 \leq x < 0.5 \\ -1 & \text{for } 0.5 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

and the d dimensional oriented wavelet functions

$$\psi^o(x) = \prod_{i=1}^{d} \psi^{o_i}(x_i)$$

where $x \in \mathbb{R}^d$, $o \in \{H, L\}^d$ and $\exists\, i, o_i = H$. Indeed, one can verify that an orthonormal basis of W_s is given by

$$\left\{\psi^o_{s,k}(x) = 2^{-sd/2}\psi^o(2^{-s}x - k)\right\}_{k\in\mathbb{Z},\, \exists i\ o_i=H}.$$

where $\psi^o_{s,k}$ is the wavelet function of orientation o rescaled by 2^s and then translated by $k2^s$. Note that functions ψ^L and ψ^H act respectively as low and high pass filters. Their combination yields oriented high pass filters, e.g., for $d = 2$, there exist three wavelet functions ψ^{HL}, ψ^{LH} and ψ^{HH}, that express details of the signal along vertical, horizontal and diagonal orientations (respectively).

We can thus decompose any function f in V_s in the two following ways:

$$\begin{aligned} f &= \sum_k a_{s,k}\phi_{s,k} \\ &= \sum_k a_{s+1,k}\phi_{s+1,k} + \sum_{o,k} d^o_{s+1,k}\psi^o_{s+1,k} \end{aligned}$$

As $\sum_k a_{s+1,k}\phi_{s+1,k}$ belongs by construction to the space V_{s+1}, we can further decompose it into a projection onto V_{s+2} and a projection onto W_{s+2}. Repeating this scheme up to scale S leads to the following multiscale decomposition of f:

$$f = \sum_k a_{S,k}\phi_{S,k} + \sum_{s'=s+1}^{S}\sum_{o,k} d^o_{s',k}\psi^o_{s',k}.$$

The classical wavelet construction is concluded by letting s go to $-\infty$, enabling one to decompose any function in such bases. Finally, we observe in particular that any function $\phi_{s',k}$ or $\psi^o_{s',k}$ with $s' \geq s$ is in V_s and therefore can be decomposed as a linear combination of $\phi_{s,k'}$:

$$\phi_{s',k} = \sum_{k'} \gamma_{s',k,s,k'}\phi_{s,k'} \tag{5}$$

$$\psi^o_{s',k} = \sum_{k'} \gamma^o_{s',k,s,k'}\phi_{s,k'} \tag{6}$$

where $\gamma_{s',k,s,k'}$ and $\gamma^o_{s',k,s,k'}$ are some fixed real numbers.

More importantly, going from the decomposition in V_s to the decomposition in the spaces V_S and $(W_{s'})_{s<s'\leq S}$ corresponds to a change of basis and thus to a discrete operation going from the coefficients $(a_{s,k})_k$ to the coefficients $(a_{S,k})_k$ and $(d^o_{s',k})_{s<s'\leq S,k,o}$. This transformation is called the Forward Wavelet Transform (FWT) and its inverse the Inverse Wavelet Transform (IWT). Both can be computed directly on the coefficients without relying on the continuous basis functions.

4.1.2 Haar Wavelet Transform of functions defined on grids

In our algorithm, rather that using the continuous Haar wavelet decomposition, we will employ the related discrete Haar wavelet decomposition on a d dimensional grid $[\![0, K_1]\!] \times ... \times [\![0, K_d]\!]$.

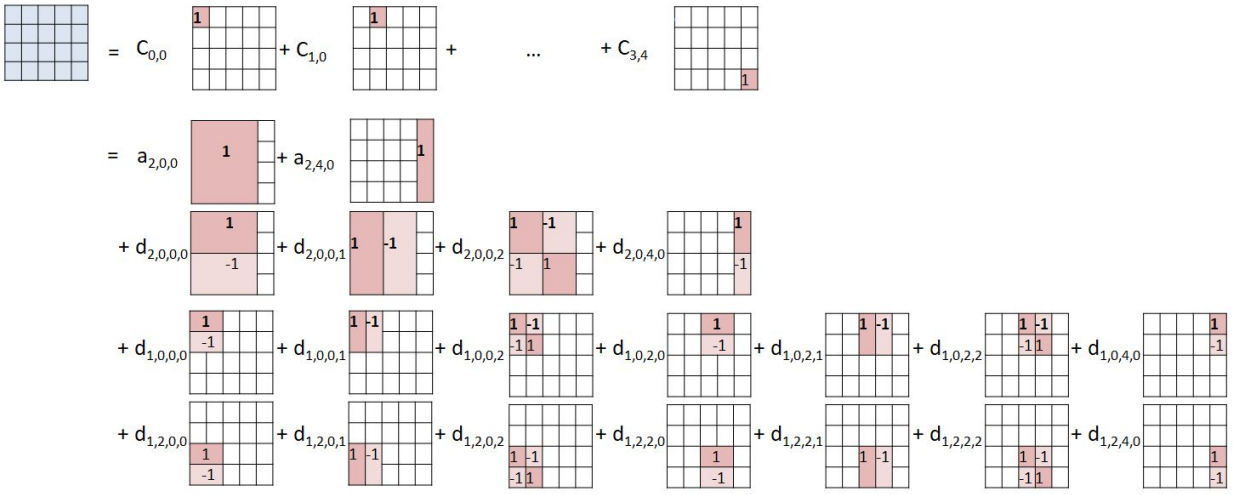

Fig. 1 Decomposition of a 4 by 5 grid according to two bases: classical decomposition into unit functions (top row) and Haar Wavelet decomposition (four bottom rows). Both decompositions yield 20 coefficients. The letters a and d refer to approximation and detail coefficients, respectively. Subscripts indicate the scale of the coefficient, the x-y position of the related wavelet function, and its orientation (in case of detail coefficients). In the grids, empty white cells denote null values

To any discrete function g on the grid, one can associate the continuous function f of V_0 defined by

$$f = \sum_{k \in [\![0,K_1]\!] \times \ldots \times [\![0,K_d]\!]} g[k]\phi_{0,k},$$

where the $(g[k])_k$ are interpreted as the approximation coefficients $(a_{0,k})_k$ at scale 0 of function f. By construction, f is a piecewise constant function on the related continuous grid. Using the FWT algorithm up to scale S, this function can be decomposed as follows:

$$f = \sum_k a_{S,k}\phi_{S,k} + \sum_{s'=s+1}^{S} \sum_{o,k} d^o_{s',k}\psi^o_{s',k}.$$

One verifies that $a_{S,k} = 0$ when $k_i < 0$ or $k_i 2^S > K_i$ and $d_{s',k} = 0$ when $k_i < 0$ or $k_i 2^{s'} > K_i$. Thus, these sums have only a finite number of coefficients. Further, when the K_i are powers of 2, i.e. $K_i = 2^{\zeta_i}$, we impose that the decomposition cannot exceed a maximum scale $S_{max} = \min_i(\zeta_i)$, so that the scaling function support remains within $\prod_i [0, K_i]$. Therefore, when $S \leq S_{max}$, the previous equation reduces to:

$$f = \sum_{0 \leq k_i < 2^{\zeta_i} - S} a_{S,k}\phi_{S,k} + \sum_{s'=s+1}^{S} \sum_{o,0 \leq k_i < 2^{\zeta_i - s'}} d^o_{s',k}\psi^o_{s',k},$$

which corresponds exactly to an orthonormal change of basis. From the coefficients $(g[k])_k$, one can compute the coefficients $(a_{S,k})_{0 \leq k < 2^{S_{max}-S}}$ and $(d^o_{s',k})_{s+1 \leq s' \leq S, 0 < k < 2^{S_{max}-s'}, o}$ with the FWT algorithm and perform the inverse operation with the IWT algorithm.

Note that the matrices M_{FWT} and M_{IWT} associated to these transformations are related to one another. We have obviously $M_{FWT} * M_{IWT} = Id$ and, as these are orthonormal transforms, we also verify that $M^T_{IWT} = M_{FWT}$. Furthermore, there exist fast implementations of these transforms that have a linear complexity with respect to the number of coefficients (Mallat, 2008).

In this work, we have implemented the FWT using the fast lifting scheme described in Mallat (1989) (Chapter 7.8). This scheme is strictly equivalent to the previous description when the grid size is a power of 2, with the difference that it handles non-dyadic grids through improved computations at the boundaries and that FWT and IWT remain orthogonal transforms. Management of the boundaries is illustrated in Figure 1, where the functions that describe approximations and details in the right border of the grid have a smaller support.

4.1.3 Preservation of the RKHS structure of the velocity fields

The previous subsection transposed the classical Haar description of continuous functions to discrete functions defined on a grid g. To implement a coarse-to-fine approach for atlas estimation, we apply the discrete Haar decomposition to grids of momentum vectors as defined in the LDDMM framework.

We recall that in the algorithm of Durrleman et al (2012), the initial geodesic vector field for a given subject i is defined as a finite linear combination of identical Gaussian kernels that are evaluated at an initial set of points:

$$v_{0,i}(x) = \sum_k K_g(x, c_{k,i}(0))\alpha_{k,i}(0)$$

where K_g is a Gaussian kernel of width σ_g and $\alpha_{k,i}(0)$ is a vector attached to $c_{k,i}(0)$. The vector field keeps this structure along time:

$$v_{t,i}(x) = \sum_k K_g(x, c_{k,i}(t))\alpha_{k,i}(t).$$

In our scheme, we set the initial points $(c_{k,i}(0))_k$ on a grid so that the first equation becomes:

$$v_{0,i}(x) = \sum_{k\in[\![0,K_1]\!]\times\ldots\times[\![0,K_d]\!]} K_g(x, c_k(0))\alpha_{k,i}(0)$$

where $c_k(0) = k$.

As the deformation field is entirely characterized by $v_{0,i}$, we do not modify the parametrization of $v_{t,i}$. The initial vector field defined at each point of the grid $[\![0, K_1]\!] \times \ldots \times [\![0, K_d]\!]$ is associated to the piecewise constant function

$$f_i(x) = \sum_{k\in[\![0,K_1]\!]\times\ldots\times[\![0,K_d]\!]} \phi_{0,k}(x)\alpha_{k,i}(0),$$

obtained by replacing $K_g(x, c_k(0))$ by $\phi_{0,k}$, which is an invertible transform. When applying the Haar transform independently to each component of $\alpha_{k,i}(0)$, we implicitly use the decomposition

$$f_i = \sum_k a_{S,k,i}\phi_{S,k} + \sum_{s'=s+1}^{S}\sum_{o,k} d^o_{s',k,i}\psi^o_{s',k}$$

and thus equivalently

$$v_{0,i} = \sum_k a_{S,k,i}\widetilde{\phi}_{S,k} + \sum_{s'=s+1}^{S}\sum_{o,k} d^o_{s',k,i}\widetilde{\psi}^o_{s',k}$$

where $\tilde{\phi}_{s,k}$ (respectively $\tilde{\psi}^o_{s,k}$) is a function defined by replacing $\phi_{0,k'}$ by $K_g(\cdot, c_{k'}(0))$ in the decomposition of $\phi_{s,k}$ (resp. of $\psi^o_{s,k}$) in Equation 5 (resp. Equation 6).

Even if we define $v_{0,i}$ through the, now vectorial, wavelet coefficients $(a_{S,k,i})_k$ and $(d^o_{s,k,i})_{1\le s\le S,o,k}$, the initial velocity field remains a combination of the original $K_g(x, c_k(0))$ and thus we are still in the setting of Durrleman et al (2012). Note that our Haar parametrization is only properly defined for $t = 0$ as the initial grid structure is deformed under the action of the diffeomorphism $\phi_{t,i}$ when the time evolves. Thus, in our optimization, we will use the Haar parametrization to update the initial velocity fields but perform all the other computations using its classical representation as a linear combination of RKHS kernels.

4.2 Coarse-to-fine atlas estimation

4.2.1 Reparametrization of the initial velocity fields

The key difference between our scheme and the one of Durrleman et al (2012) is the use of the Haar parametrization in a fixed grid for the initial velocity fields. Thus, we optimize a cost function

$$E(I_0, (\beta_{0,i})_{1\le i\le N}) = \sum_{i=1}^{N}\left(\frac{\|I_i - I_0\circ\Phi^{-1}_{1,i}\|^2}{2\sigma^2} + \|v_{0,i}\|^2_V\right),$$

where $\beta_{0,i}$ is a set of wavelet coefficients $(a_{S,k,i})_k$ and $(d^o_{s,k,i})_{s,k,o}$, related to the momentum vectors $\alpha_{0,i}$ by $\alpha_{0,i} = IWT(\beta_{0,i}) = M_{IWT}\beta_{0,i}$. This implies a relationship between the gradient $\nabla_{\beta_{0,i}}E$ of the cost function with respect to the wavelet coefficients $\beta_{0,i}$ and the gradient $\nabla_{\alpha_{0,i}}E$ with respect to the momenta $\alpha_{0,i}$:

$$\begin{aligned}\nabla_{\beta_{0,i}}E &= M^T_{IWT}\nabla_{\alpha_{0,i}}E \\ &= M_{FWT}\nabla_{\alpha_{0,i}}E = FWT(\nabla_{\alpha_{0,i}}E)\end{aligned}$$

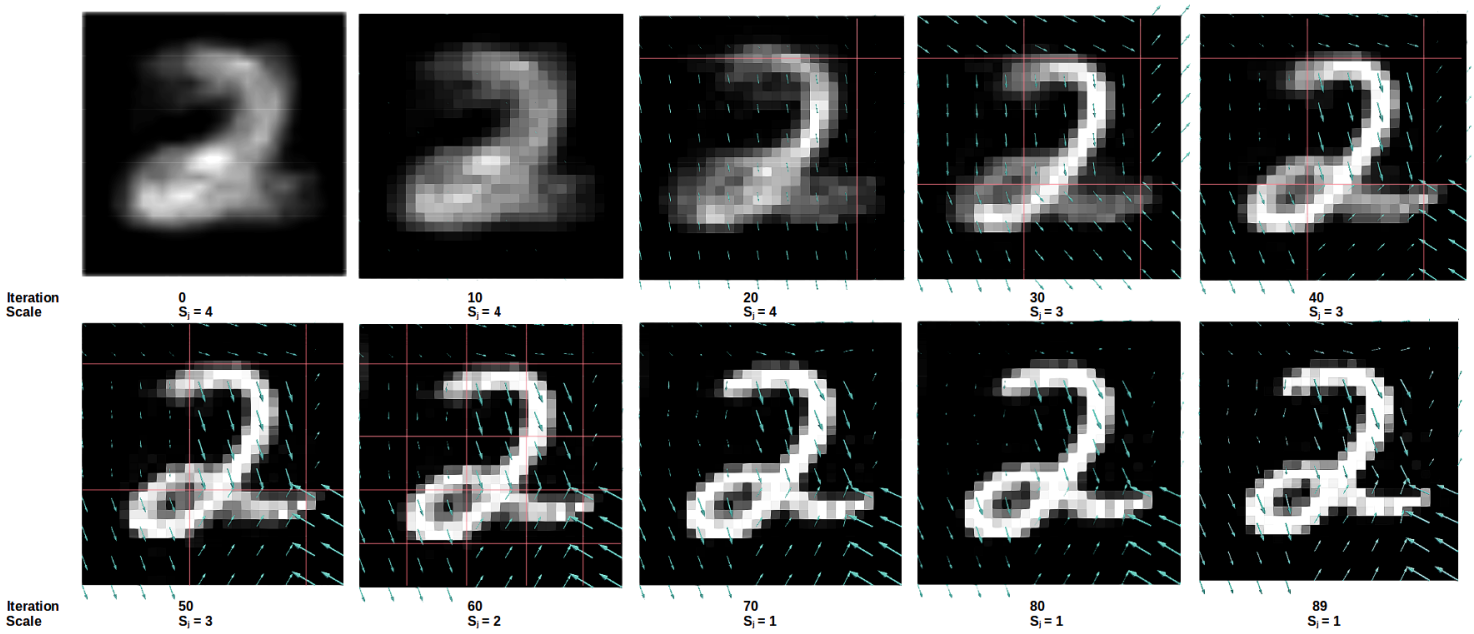

Fig. 2 Coarse-to-fine atlas estimation algorithm applied to handwritten digits with $\sigma_g = 3$ (corresponding to 100 control points). The estimated template image is displayed every 10 iterations until convergence. Blue arrows indicate the estimated momentum vectors that wrap the template to the image of the first subject. Magnitude of the vectors was multiplied by 1.5 for visualization purposes. Coarse-to-fine steps were performed at iterations 26, 56 and 64. Red lines delineate the support of the Haar wavelet functions at each scale S_j. Control points belonging to the same area are updated with identical momenta values. At scale $S_j = 1$, the support of these functions is of size 1x1 control points and the velocity fields are not constrained anymore. Note that as the grid of control points is not a power of 2, the support of the wavelet functions located at the borders is reduced.

where we have used the fact that $M_{IWT}^T = M_{FWT}$ because we have an orthonormal transform. The gradient $\nabla_{\beta_{0,i}} E$ can thereby be computed for almost the same cost as the one of this latter gradient. If we use the same gradient descent algorithm as in the original algorithm but in the wavelet domain, we obtain exactly the same results as the original algorithm.

To obtain different and hopefully better results, we enforce some constraints on the wavelet coefficients of the initial velocity fields. Namely, we use a coarse-to-fine initialization strategy by optimizing first the initial velocity fields whose wavelets coefficients are null at the finest scales and adding progressively these fine scale coefficients. In the following, we describe in detail our procedure which is summarized in Algorithm 1 and illustrated on a simple example in Figure 2.

4.2.2 Coarse-to-fine optimization

The coarse-to-fine optimization can be seen as an initialization of each new scale with the optimal template-to-subject deformations of the previous coarser scale. Indeed, we have observed that this seems to decrease the probability of trapping the optimization in a local minimum linked to unrealistic deformations, especially in cases where the images are complex and the deformation grid is fine.

More precisely, at iteration j, we only optimize the wavelet coefficients whose scales are above or equal to a current decreasing scale S_j, and set the coefficients below this scale equal to 0. This can easily be done by applying FWT to $\nabla_{\alpha_{0,i}} E$ to derive the gradient with respect to $\beta_{0,i}$ and then setting to 0 all the wavelet coefficients whose scale is strictly smaller than S_j ("Finer scale silencing step" in Algorithm 1). We then update the coefficients $\beta_{0,i}$ with the modified gradient, and recover the updated $\alpha_{0,i}$ using the IWT function ("Parameters update step" of Algorithm 1). If we iterate without modifying the current scale, we optimize the cost function in a subspace of functions that are simpler than in the original algorithm, the wavelet transform scale limitation acting as a regularizer.

As we want to optimize on the full set of functions defined by the momentum vectors, we progressively decrease the current scale S_j. We propose to decrease the scale when we are close to convergence at the current scale ("Scale refinement step" of Algorithm 1) until

Algorithm 1 Coarse-to-fine atlas estimation algorithm

Input
Set of images $(I_i)_{1\leq i\leq N}$ of dimension d, template image I_0, geometric kernel width σ_g, trade-off regularity/fidelity-to-data σ, initial step size h
Initialization
$j \leftarrow 0$
Regular grid of control points c_0 with spacing σ_g
Template image intensities $I_0(j) = I_0$
Momentum vectors $\alpha_{0,i}(j) \leftarrow 0$ for each subject i
Compute $\beta_{0,i}(j) = FWT(\alpha_{0,i}(j))$ for each subject i
Current scale S_j = maximum scale of $\beta_{0,0}(j)$
repeat
    $j \leftarrow j + 1$
    Compute the gradients $\nabla_{\alpha_{0,i}} E$ for all subjects i and the gradient $\nabla_{I_0} E$
    for each subject i **do**
        $\nabla_{\beta_{0,i}} E = FWT(\nabla_{\alpha_{0,i}} E) = (a^i_{S_{max},k})_k \cup (d^{i,o}_{s,k})_{1\leq s\leq S_{max},k,o}$
        for each detail coefficient $d^{i,o}_{s,k}$ of $\nabla_{\beta_{0,i}E}$ **do**
            if $s < S_j$ **then** ▷ Finer scale silencing step
                $d^{i,o}_{s,k} = 0$
            end if
        end for
    end for
    $\beta_{0,i}(j) = \beta_{0,i}(j-1) - h \times \nabla_{\beta_{0,i}} E$ for all subjects i ▷ Parameter update step
    $\alpha_{0,i}(j) = IWT(\beta_{0,i}(j))$ for all subjects i
    $I_0(j) = I_0(j-1) - h \times \nabla_{I_0} E$

    Compute the total residual value R_j
    if $\frac{R_{j-1}-R_j}{R_{j-1}} < 0.01$ and $S_j > 1$ **then**
        $S_j = S_{j-1} - 1$ ▷ Scale refinement step
    end if
until Convergence
Output
Template image I_0 and momentum vectors $\alpha_{0,i}$

we reach the finest possible scale. This is measured by computing the mean residual norm over subjects at each step:

$$\Delta_j(x) = \frac{1}{N}\sum_i \|I_{j,0} \circ \Phi^{-1}_{j,i} - I_i\|_2^2, \quad (7)$$

where j denotes the current iteration.

We then compute the total residual value $R_j = \|\Delta_j\|_1$. If the residual decrease with respect to the previous iteration $j-1$ is below a threshold of 1%, we decide that the algorithm is close to convergence. In the case of Figure 2, the algorithm starts at $S_0 = 4$, performs optimization until (almost) convergence at this scale, diminishes the scale to 3 and performs the subsequent scale transitions in the same manner.

At a given scale S_j, the velocity fields are allowed to vary spatially only at scales coarser than $S_j - 1$. In other words, at scale S_j, the initial velocity field of a given subject i implicitly writes as follows:

$$v_{0,i}(x) = \sum_k a_{S_{max},k,i}\widetilde{\phi_{S_{max},k}} + \sum_{s'=S_j}^{S_{max}}\sum_{k,o} d^o_{s',k,i}\widetilde{\psi^o_{s',k}}$$

where $\widetilde{\phi_{S_{max},k}}$ and $\widetilde{\psi^o_{s',k}}$ are linear combinations of localized Gaussian kernels $K_g(x, c_k(0))$.

This procedure ensures that the momenta belonging to the same zone (as defined by the support of the wavelet functions at scale S_j) are updated with identical values. This phenomenon is illustrated in Figure 2, where red

lines delineate the support of the wavelet functions at scale S_j and thus the areas where the velocity fields are constrained to have even values.

At each coarse-to-fine step, the support of the wavelet functions defined at scale S_j is decreased by a factor of 2^d - and so is the size of the areas in which the velocity fields have even values. When the algorithm reaches scale 1, the momenta are updated independently of each other. Importantly, unlike previous approaches that represented deformations in a wavelet basis (Amit, 1994; Downie et al, 1996; Wu et al, 2000; Sun et al, 2014; Gefen et al, 2004; Torbati and Ayatollahi, 2017), when the algorithm reaches this finest scale, the momentum vectors are completely free of constraints and the parametrization of the velocity fields is equivalent to its original definition (i.e. a sum of localized small Gaussian kernels). Thus, in theory, our coarse-to-fine algorithm could reach the same solutions as the original one, but as we will see in Section 5, the coarse-to-fine numerical scheme converges to better solution.

4.2.3 Complexity

Contrary to the previous coarse-to-fine algorithms developed in the LDDMM framework, our strategy does not add any complexity to the mathematical model. The parametrization of the velocity fields remains identical to that of Durrleman et al (2012). Computation of the gradients and subsequent cost also remains identical. The only additional complexity originates from the algorithms FWT and IWT, which, following the implementation of Mallat (2008), are of linear complexity.

5 Experiments

In this section, we compare our algorithm to the original version from Durrleman et al (2012) on three different datasets. The training phase consists in atlas estimation and the test phase consists in registering the estimated template image to a set of new images. Performance of the algorithms is assessed by computing the mean percentage decrease in residuals after atlas estimation or registration, and by evaluating visually the quality of the template images.

We compare the original version of the algorithm, which is available in Deformetrica Version 4.3.0, against our modified version which includes the wavelet reparametrization of the initial velocity fields. Optimization relies on a gradient descent algorithm in which the step sizes h are first scaled by the squared norm of the gradients and then diminished by a backtracking algorithm to guaranty a descent. The following parameters are used: $\sigma = 0.1$ in the cost criterion; initial step size $h = 0.01$; convergence threshold $= 0.0001$. The initial template image for atlas estimation are given by the mean of the intensities of the training images.

5.1 Handwritten digits

In this section, we use images of the digit two extracted from the well-known United States postal database of handwritten digits (Hastie et al, 2004). The size of the images is 28 by 28 pixels. We test the original algorithm against the coarse-to-fine strategy using an experimental procedure similar to that of (Durrleman et al, 2012): atlas estimation is performed using 20 randomly-chosen training images and the estimated template image is registered to 10 randomly chosen test images with the same parameters as those used during training. This experiment is repeated five times with different training and test sets. There is no intersection between any of the training and test sets. This procedure is performed for each version of the algorithm and reproduced with different sets of parameters (number of control points and width of the Gaussian kernel). Since the five experiments are performed on independent datasets, we use paired Student t-tests to compare performance between the two algorithms. Table 1 shows the mean residual decrease achieved by the algorithms over the experiments. Figure 3 presents the five template images estimated by each algorithm with three different sets of parameters, along with the template image estimated from the first training set warped to five of the training images.

Table 1 shows that the coarse-to-fine algorithm outperforms the original version in terms of residual decrease during both atlas estimation and registration, with differences that reach significance for $\sigma_g = 2$ and $\sigma_g = 1.5$. Consistent with these results, we observe in Figure 3 that the original algorithm yields highly noisy template images, a trend which worsens when the number of control points increases, indicating overfitting. The transformation of the first template image towards the training images yields reconstructed images that are close to the original ones for $\sigma_g =$

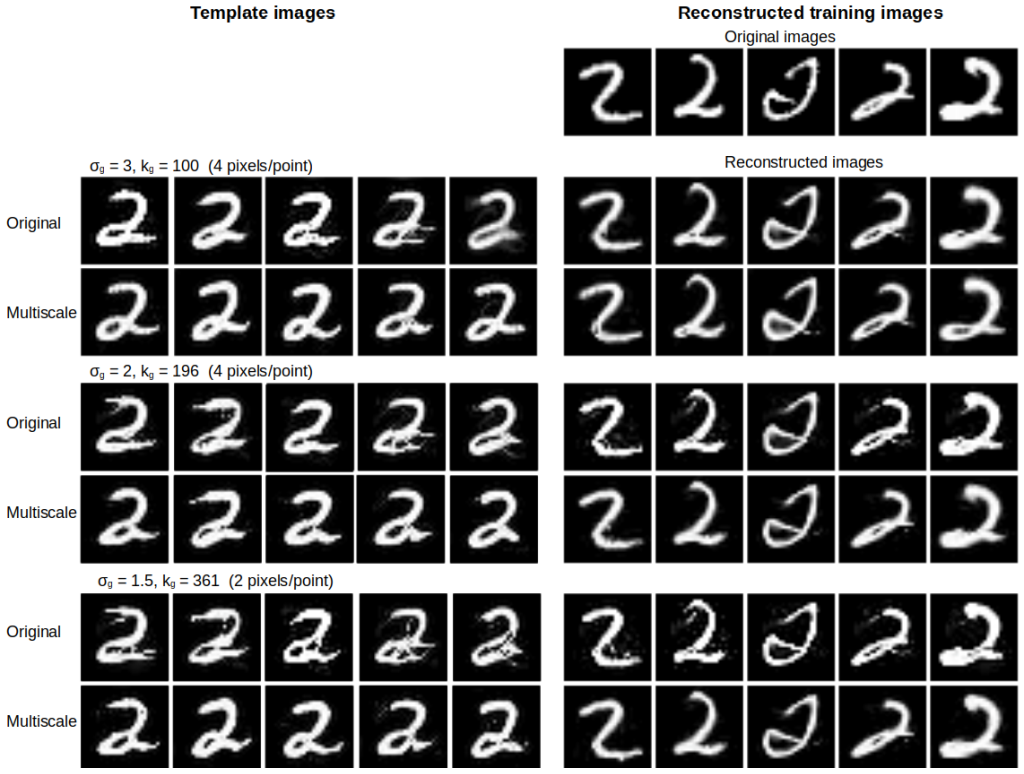

Fig. 3 Estimation of the template image by the original (top rows) and coarse-to-fine (bottom rows) algorithms on the dataset of handwritten digits with 3 different parameters σ_g. For each experiment, five template images (estimated with non-intersecting training sets) are presented on the left, along with the template image estimated from the first training set wrapped to the first five training images on the right. σ_g and k_g indicate the width of the Gaussian kernel and the corresponding number of control points, respectively. The number of voxels per control point are indicated in parentheses.

3, but these deformations become noisy for lower values of σ_g. These observations belie the quantitative evaluation, which shows that the performance of the original algorithm increases with the number of parameters. This discrepancy demonstrates that residual decrease alone is not sufficient to evaluate the accuracy of the algorithms, as irregular vector fields might be able to match any subject at the cost of producing unnatural deformations.

Unlike the original algorithm, the coarse-to-fine procedure produces realistic template images, whose quality is preserved when decreasing σ_g. Despite the use of non-intersecting training sets, we notice that template images are quite similar to each other. Moreover, all reconstructed images are very close to the original ones. Their quality slightly increases with the number of control points: this is most evident for the third and fifth reconstructed subjects, which become more accurate for lower values of σ_g.

σ_g	k_g	Original	Coarse-to-fine
		Atlas estimation	
3	100	89.8 ± 3.1	92.1 ± 1.9
2	196	89.9 ± 1.7	$\mathbf{94.7 \pm 0.5}$
1.5	361	92.7 ± 1.7	$\mathbf{95.7 \pm 1.4}$
		Registration	
3	100	90.0 ± 6.1	92.5 ± 6.0
2	196	91.5 ± 5.2	$\mathbf{93.5 \pm 6.3}$
1.5	361	91.0 ± 5.1	$\mathbf{94.7 \pm 7.0}$

Table 1 Performance of the original and coarse-to-fine algorithms on the dataset of handwritten digits during the training and test phases. Data are mean ± standard deviation of residual decrease over five experiments (in percentage). Bold style indicates that the coarse-to-fine algorithm performed significantly better than the original one (pvalue< 0.05). σ_g: width of the Gaussian kernel; k_g: number of control points.

5.2 Artificial characters

In the previous experiment, one can remark that the algorithm's performance diverge most when using a high number of parameters. Therefore, one might simply be tempted to employ the original algorithm with a lower

number of parameters, as in Durrleman et al (2012) whose experiments were performed with 36 control points. However, datasets that present a higher amount of details and inter-subject variability may benefit from our coarse-to-fine strategy even when a lower number of control points is used. To confront our algorithm with a more difficult task, we manually designed a dataset of 30 characters (displayed in Appendix A). The size of the images is 28 by 28 pixels.

We compare our algorithm to the original version using cross-validation: the dataset is randomly split into a training set (24 images) and a test set (6 images). Each algorithm estimates independently a template image from the training set, and then registers the template to each image in the test set with the same parameters as those used during training. This procedure is repeated five times, and reproduced with different parameters of the algorithms (number of control points and width of the Gaussian kernel). No statistical tests are performed because of the overlap between the training sets and between the test sets. Table 2 displays the mean residual decrease during atlas estimation and registration and Figure 4 presents the five template images estimated by each algorithm with three different sets of parameters, along with five reconstructed training images.

σ_g	k_g	Original	Coarse-to-fine
		Atlas estimation	
5	36	86.0 ± 4.5	94.5 ± 0.8
4	49	83.6 ± 3.8	95.6 ± 0.4
3	100	81.6 ± 4.5	96.7 ± 0.5
		Registration	
5	36	90.2 ± 6.2	94.0 ± 6.3
4	49	89.3 ± 6.0	95.4 ± 4.2
3	100	86.6 ± 7.5	96.4 ± 3.8

Table 2 Performance of the original and coarse-to-fine algorithm on the dataset of artificial characters during the training and test phases. Data are mean $\pm$ standard deviation of residual decrease over five folds of cross-validation (in percentage). σ_g: width of the Gaussian kernel; k_g: number of control points.

Table 2 shows that the coarse-to-fine algorithm reaches higher performance than the original algorithm. The performance of the coarse-to-fine strategy increases with the number of control points during train and test, while the original algorithm demonstrates the opposite trend. These results are supported by the qualitative evaluation of the template images. In Figure 4, for $\sigma_g = 5$, the two algorithms generate template images that present discrete but noticeable differences. In case of the original version, the arms and legs of the characters appear slightly fuzzier, and the second template image is noisy. This observation is confirmed by comparing the reconstructed images: those yielded by the original algorithm are blurry (and even erroneous in case of the third subject), while the template and reconstructed images yielded by the coarse-to-fine algorithm seem sharp and accurate. As with the dataset of handwritten digits, the quality of the template images estimated by the original algorithm deteriorates when increasing the number of control points: the images become fuzzier and display more erroneous features such as additional arms below the character's neck, present in four of the template images estimated for $\sigma_g = 4$ and $\sigma_g = 3$. The morphology of all but the fifth reconstructed characters is also completely erroneous. In contrast, the coarse-to-fine algorithm is able to produce very stable, sharp and correct template images for all parameters. Similar observations can be made regarding the transformation of the template image towards the five training images: the coarse-to-fine strategy succeeds in generating images that are nearly identical to the original ones.

These differences have a simple explanation: the original version simultaneously estimates the overall shape of the characters and details such as the location and orientation of the arms and legs, making it more dependent on the initial template image and leading to the selection of erroneous features, while the coarse-to-fine strategy first focuses on estimating the characters main features, which are then refined during the finer scales. This phenomenon is illustrated by movies that show the evolution of the estimated templates across iterations, available at the first author's webpage[2].

This experiment shows that on a dataset with high variability, the original algorithm is unable to estimate templates that are satisfying with respect to quantitative and qualitative criteria. The coarse-to-fine algorithm outperforms the former in both regards regardless of the chosen number of parameters.

[2] https://fleurgaudfernau.github.io/Multiscale_atlas_estimation/

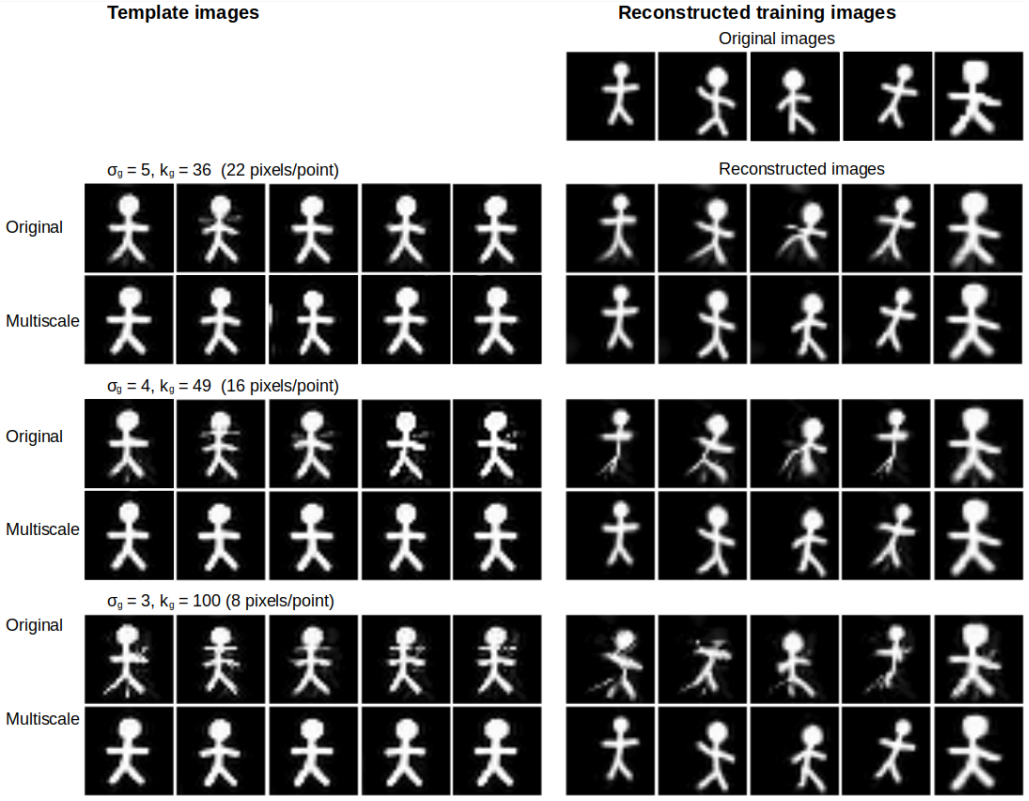

Fig. 4 Estimation of the template image by the original (top rows) and coarse-to-fine (bottom rows) algorithms on the dataset of artificial characters with three different parameters σ_g. For each experiment, five estimated template images (for each fold of cross-validation) are presented on the left, along with the template image estimated from the first training set wrapped to the first five training images on the right. σ_g indicates the width of the Gaussian kernel, k_g indicates the corresponding number of control points used, and the number of pixels per control point are indicated in parentheses.

5.3 Fetal brain MRIs

σ_g	k_g	Original	Coarse-to-fine
		Atlas estimation	
10	1,320	50.2 ± 3.4	52.3 ± 3.4
7	4,050	56.4 ± 3.7	62.2 ± 0.9
5	10,080	61.5 ± 4.0	67.5 ± 1.6
		Registration	
10	1,320	58.9 ± 6.9	60.1 ± 6.7
7	4,050	65.0 ± 6.2	68.9 ± 5.7
5	10,080	69.6 ± 5.0	74.4 ± 5.2

Table 3 Performance of the original and coarse-to-fine algorithm on the dataset of fetal brains MRIs during the training and test phases. Data are mean $\pm$ standard deviation of residual decrease over five folds of cross-validation (in percentage). σ_g: width of the Gaussian kernel; k_g: number of control points.

To evaluate the performance of our coarse-to-fine approach on a dataset of clinical images, we select 30 fetal brain MRIs with agenesis of the corpus callosum acquired in Hopital Trousseau, France. Gestational ages are comprised between 32 and 34 weeks of gestation (mean = 32.9 ± 0.6). A more detailed presentation of the dataset is available in Gaudfernau et al (2021).

Agenesis of the corpus callosum is a developmental anomaly characterized by the total or partial absence of the corpus callosum. It is often associated to anatomical features such as widening of the lateral ventricles. Atlas estimation can help better understand congenital anomalies by providing an insight into how these anatomical characteristics vary together (Gaudfernau et al, 2021). However, as abnormal fetal brains may present a wide range of defects, this makes atlas estimation more difficult and prone to errors than with datasets of healthy fetuses. Thus, it is crucial to develop algorithms that are capable of estimating realistic templates on both healthy and abnormal subjects.

The brain MRIs are preprocessed, volume reconstructed and rigidly aligned according to the procedure described in Gaudfernau et al

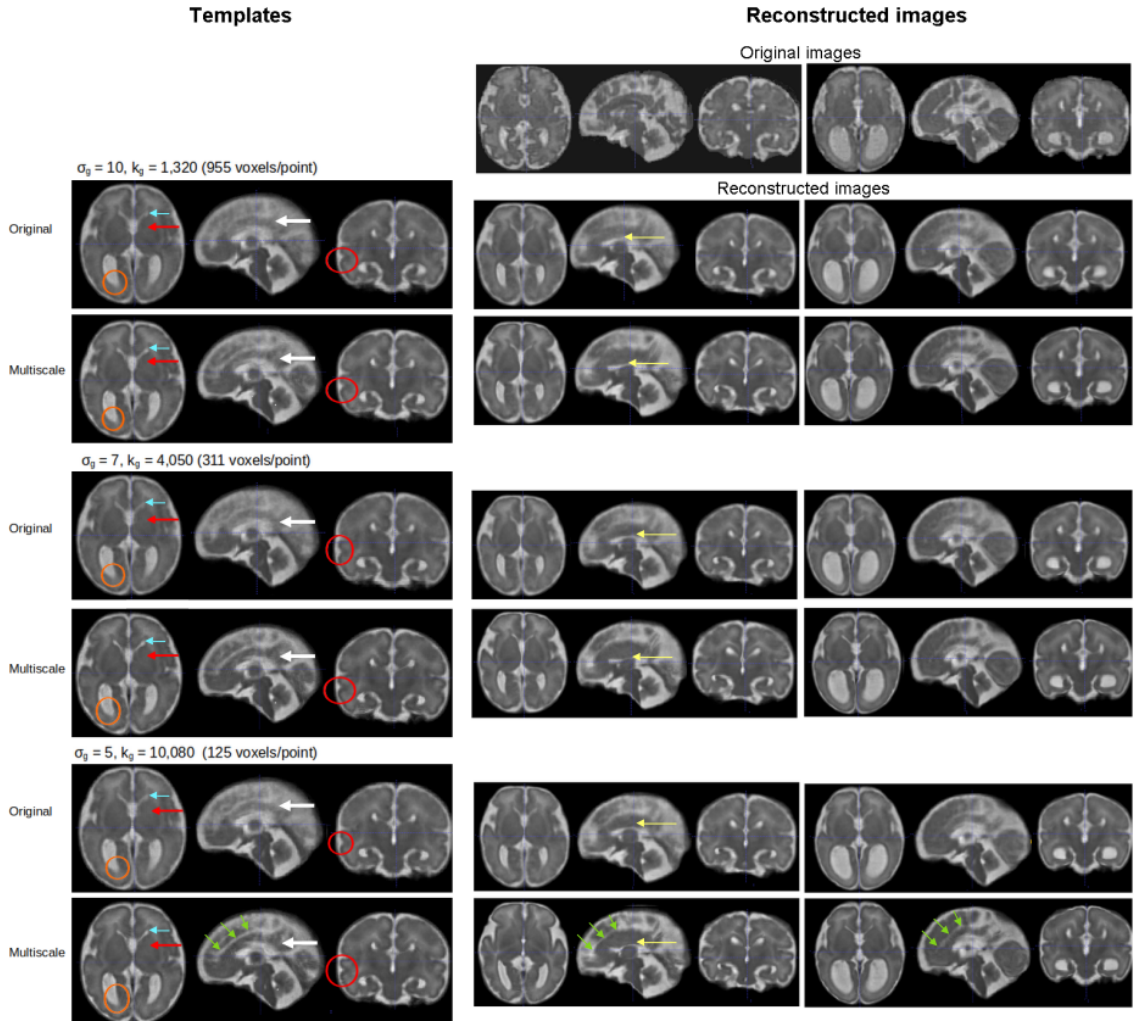

Fig. 5 Estimation of the template image by the original and coarse-to-fine algorithms on the dataset of fetal brain MRIs with 3 different parameters σ_g. σ_g and k_g indicate the width of the Gaussian kernel and the corresponding number of control points, respectively. The number of voxels per control point are indicated in parentheses. For each experiment, the estimated template image from the first fold of cross-validation is presented in axial, sagittal and coronal views in the left column. The first row displays the first two original training images; the middle and right columns display the corresponding reconstructed images (i.e. the template image warped to the two training images). Salient differences between the algorithms are indicated by specific markers: white arrows indicate the cingulum bundle, blue arrows the left frontal horn, red arrows the left basal ganglia, orange circles the posterior part of the right lateral ventricle, red circles the right superior temporal sulcus, yellow arrows the theoretical location of the corpus callosum and green arrows the cingulate sulcus.

(2021). The final images have size 105x100x120 voxels. Cross-validation is performed in the same manner as in section 5.2, with 24 images used for atlas estimation and 6 images used for testing. Figure 5 presents one example of the five estimated template images during cross validation for different parameters σ_g, and Table 3 displays the mean residual decrease yielded by the algorithms.

Table 3 shows the importance of setting a high number of parameters when performing atlas estimation on clinical images with high variability: the performance of both algorithms in train and test rises with the number of control points, which is reflected by the increasing sharpness of the template images in Figure 5.

During training, the mean residual decrease of the coarse-to-fine algorithm is superior to that of the original algorithm for all parametrizations. The estimated coarse-to-fine atlases are also more stable, as indicated by lower values of standard deviation during training. Registration results demonstrate that the coarse-to-fine algorithm has a higher ability to generalize to new images.

In terms of template quality, the original and coarse-to-fine algorithms yield distinct

results. The brain images estimated with the latter display sharper features and enhanced contrast between structures, especially for higher values of σ_g. The template images estimated by the original algorithm display more fuzzy areas, especially in the cingulum bundle (white arrows), the frontal horns (blue arrows) and the posterior area of the lateral ventricles (orange circles).

On the contrary, these structures are more clearly delineated in the template images generated by the coarse-to-fine algorithm. In addition, this method leads to more pronounced gyration patterns (e.g. the superior temporal sulcus, see red circles) and more visible subcortical brain structures such as the basal ganglia (red arrows). Interestingly, the most salient differences between the estimated templates are located at the medial surface of the brain (sagittal view), which presents high variability between subjects: the surface estimated by the original algorithm is close to the mean of the image intensities, while the coarse-to-fine strategy produces a realistic anatomy, with a visible cingulate sulcus (green arrows).

Some differences can also been noted between the reconstructed images yielded by the algorithms. In addition to increased sharpness, the coarse-to-fine template image warped to the subjects reveals more abnormal features, such as abnormally shaped corpus callosum in the case of the first subject (yellow arrows, middle column).

All together, these results indicate that our coarse-to-fine algorithm can be successfully applied to real-world, complex clinical data.

6 Discussion

In this paper, we introduced a novel, wavelet-based reparametrization of the initial velocity fields in the LDDMM framework. Taking advantage of the hierarchical property of the wavelet decomposition, we implemented a coarse-to-fine optimization for atlas estimation. This transfer of information from one scale to the next ensures smarter initialization of the deformations at each level, leading the algorithm to favor smoother solutions while avoiding unrealistic local minima. Contrary to previous coarse-to-fine algorithms introduced in the LDDMM framework (Sommer et al, 2012; Gris et al, 2016; Modin et al, 2018; Miller et al, 2020), our approach adds very little complexity to the mathematical model. Specifically, the reparametrization of the velocity fields can be seen as an additional layer of spatial regularization, which preserves the RKHS structure of the vector fields and the efficient numerical scheme used to compute the gradients.

We performed experiments on three datasets of increasing difficulty. Compared to the original algorithm, the coarse-to-fine algorithm yields higher quality templates with better stability, that are able to generalize to unseen images. Not only does our strategy produce images that have a realistic anatomy, but it leads to enhanced preservation of anatomical details, including unusual or abnormal ones. This makes it particularly appropriate for tasks involving high inter-subject variability, specifically clinical images.

Some limitations of this algorithm have to be highlighted. Unlike approaches based on a mixture of kernels, our deformation model relies on a single Gaussian kernel. While this has the advantage of introducing no additional parameters, the results of our algorithm depend on the choice of the kernel width σ_g - but to a lesser extent than the original version. Some methodological improvements could be made to the FWT algorithm. In particular, the border treatment of non dyadic grids define scaling and wavelet functions of small support in some image borders. As illustrated in Figure 2, this creates areas of uneven size at coarse scales, and therefore vector fields with locally uneven smoothness constraints. This could pose a problem if one were to exploit images with borders showing high variability. To avoid this, one could replace the Haar wavelet by smoother wavelet functions such as the Daubechies wavelet. However, since image borders often contain little information, this issue is unlikely to arise.

The properties of the wavelet transform offer interesting avenues to explore. Notably, the wavelet functions $\phi_{s,k}$ and $\psi^o_{s,k}$ are normalized depending on their scale s. By modifying the weights attributed to the fine and coarse scales in this normalization, one can change the relative importance attributed to high and low frequency coefficients during optimization. It would be interesting to refine our wavelet-based spatial regularizer in this way and observe how it impacts the results.

The simplicity of our algorithm makes it easy to implement, which opens up interesting perspectives. The coarse-to-fine strategy could

be applied to other types of atlases such as spatio-temporal ones (Debavelaere et al, 2020), or to other statistical frameworks such as the Bayesian Mixed Effect Model (Allassonnière et al, 2015). In regard to the latter point, a significant advance would be to integrate our reparametrization of the velocity fields into this Bayesian framework by introducing priors on the wavelet coefficients, in the spirit of Downie et al (1996) who decomposed deformations into a Haar wavelet basis and modelled the coefficients as independent random variables with a mixture distribution. Moreover, our approach could be integrated into different mathematical settings, such as Modin et al (2018); Miller et al (2020), where the estimation of the vector fields at each scale of the composition could benefit from our efficient coarse-to-fine optimization.

In the future, several improvements will be explored, starting with the addition of a multiscale flavor to our algorithm, in line with strategies based on simultaneously coexisting flows (Risser et al, 2011). Specifically, we will work on constraining the velocity fields to be unevenly smooth according to the amount of information contained in the images. Unlike (Risser et al, 2011), since our algorithm preserves the RKHS structure of the velocity fields and thus the definition of the norm, this would not add complexity to the mathematical model nor to the optimization.

In addition, we will focus on developing a dual coarse-to-fine strategy, by applying a hierarchical representation to the images, as already attempted in other mathematical models, e.g. with B-spline deformations (Rueckert et al, 1999; Loeckx et al, 2007; Staring et al, 2009) and in the hyperelasticity framework (Debroux et al, 2021). Alternating both coarse-to-fine strategies would very likely provide template images of even higher quality.

Declarations

Funding. This work was partly funded by the third author's chair in the PRAIRIE institute funded by the French national agency ANR as part of the "Investissements d'avenir" programme under the reference ANR-19- P3IA-0001.

Conflict of interest. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval. This work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of human subjects. The experiments conducted on the dataset of fetal brain MRI has been granted ethics approval by the Institutional Review Board of the Comité d'éthique de la recherche en imagerie médicale (CERIM) under the reference CRM-2112-215.

Data. The dataset of artificial characters can be downloaded at the first author's webpage[3].

Code availability. Our code is publicly available in a public Git repository[4].

[3] https://fleurgaudfernau.github.io/Multiscale_atlas_estimation/

[4] https://github.com/fleurgaudfernau/Deformetrica_multiscale/

Appendix A Dataset of artificial characters

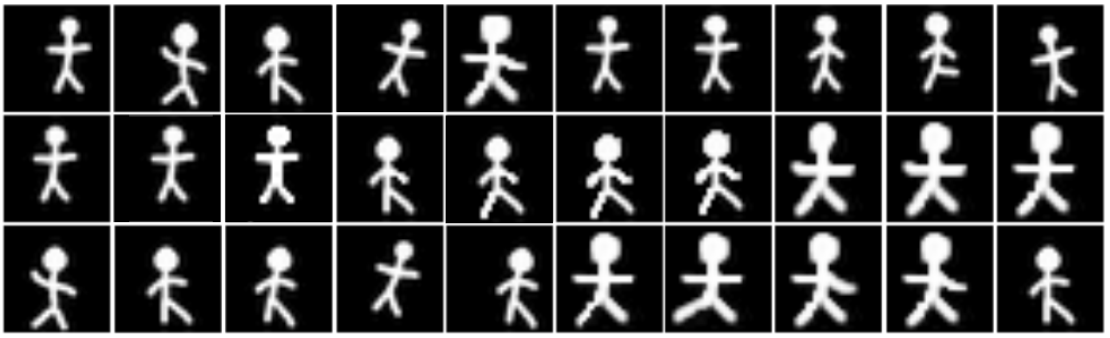

Fig. A1 Dataset of manually designed characters.

References

Allassonnière S, Durrleman S, Kuhn E (2015) Bayesian mixed effect atlas estimation with a diffeomorphic deformation model. SIAM Journal on Imaging Sciences 8:1367–1395. https://doi.org/10.1137/140971762

Amit Y (1994) A nonlinear variational problem for image matching. Siam Journal on Scientific Computing 15. https://doi.org/10.1137/0915014

Athavale P, Xu R, Radau P, et al (2015) Multiscale properties of weighted total variation flow with applications to denoising and registration. Medical Image Analysis 23. https://doi.org/10.1016/j.media.2015.04.013

Bruveris M, Risser L, Vialard FX (2011) Mixture of kernels and iterated semidirect product of diffeomorphisms groups. SIAM Journal on Multiscale Modeling and Simulation 10. https://doi.org/10.1137/110846324

Christensen G, Rabbitt R, Miller M (1996) Deformable template using large deformation kinematics. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society 5:1435–47. https://doi.org/10.1109/83.536892

Debavelaere V, Durrleman S, Allassonnière S (2020) Learning the clustering of longitudinal shape data sets into a mixture of independent or branching trajectories. International Journal of Computer Vision 128. https://doi.org/10.1007/s11263-020-01337-8

Debroux N, Le Guyader C, Vese LA (2021) Multiscale Registration. In: Scale Space and Variational Methods in Computer Vision, Lecture Notes in Computer Science, vol 12679. Springer International Publishing, p 115–127, https://doi.org/10.1007/978-3-030-75549-2_10, URL https://hal.uca.fr/hal-03470756

Downie T, Shepstone L, Silverman B (1996) A wavelet based approach to deformable templates. In: Proceedings in Image Fusion and Shape Variability Techniques

Durrleman S, Allassonnière S, Joshi S (2012) Sparse adaptive parameterization of variability in image ensembles. International Journal of Computer Vision 101:1–23. https://doi.org/10.1007/s11263-012-0556-1

Gaudfernau F, Blondiaux E, Allassonière S (2021) Analysis of the Anatomical Variability of Fetal Brains with Corpus Callosum Agenesis. In: Sudre CH, Licandro R, Baumgartner C, et al (eds) MICCAI 2021 - Perinatal, Preterm and Paediatric Image Analysis Workshop, LNCS - Lecture Notes in Computer Science, vol 12959. Springer, Strasbourg, France, https://doi.org/10.1007/978-3-030-87735-4_26

Gefen S, Tretiak O, Bertrand L, et al (2004) Surface alignment of an elastic body using a multiresolution wavelet representation. IEEE transactions on bio-medical engineering 51:1230–41. https://doi.org/10.1109/TBME.2004.827258

Gholipour A, et al (2017) A normative spatiotemporal mri atlas of the fetal brain for automatic segmentation and analysis of early brain growth. Scientific Reports 7. https://doi.org/10.1038/s41598-017-00525-w

Grenander U, Miller M (1998) Computational anatomy: An emerging discipline. Quarterly of Applied Mathematics LVI:617–694. https://doi.org/10.1090/qam/1668732

Gris B, Durrleman S, Trouvé A (2016) A sub-riemannian modular framework for diffeomorphism-based analysis of shape ensembles. SIAM Journal on Imaging Sciences 11. https://doi.org/10.1137/16M1076733

Hastie T, Tibshirani R, Friedman J, et al (2004) The elements of statistical learning: Data mining, inference, and prediction. Math Intell 27:83–85. https://doi.org/10.1007/BF02985802

Kaplan J, Donoho D (2021) The morphlet transform: A multiscale representation for diffeomorphisms, phD thesis

Lam KC, Ng TC, Lui LM (2017) Multiscale representation of deformation via beltrami coefficients. Multiscale Modeling & Simulation 15:864–891. https://doi.org/10.1137/16M1056614

Loeckx D, Slagmolen P, Maes F, et al (2007) Nonrigid image registration using conditional mutual information. Information processing in medical imaging : proceedings of the conference 20:725–37. https://doi.org/10.1109/TMI.2009.2021843

Mallat S (1989) Theory for multi-resolution signal decomposition - the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence - PAMI 11

Mallat S (2008) A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, 3rd edn. Academic Press, Inc., USA

Miller M, Trouvé A, Younes L (2002) On the metrics and euler-lagrange equations of computational anatomy. Annual review of biomedical engineering 4:375–405. https://doi.org/10.1146/annurev.bioeng.4.092101.125733

Miller M, Trouvé A, Younes L (2006) Geodesic shooting for computational anatomy. Journal of mathematical imaging and vision 24:209–228. https://doi.org/10.1007/s10851-005-3624-0

Miller MI, Tward DJ, Trouvé A (2020) Coarse-to-fine hamiltonian dynamics of hierarchical flows in computational anatomy. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 3760–3765, https://doi.org/10.1109/CVPRW50498.2020.00438

Mk A, Freeden W, Mayer C (2003) Multiscale deformation analysis by cauchy-navier wavelets. Journal of Applied Mathematics 2003. https://doi.org/10.1155/S1110757X03206033

Modersitzki J (2008) FAIR: Flexible Algorithms for Image Registration. https://doi.org/10.1137/1.9780898718843

Modin K, Nachman A, Rondi L (2018) A multiscale theory for image registration and nonlinear inverse problems. Advances in Mathematics 346. https://doi.org/10.1016/j.aim.2019.02.014

Niethammer M, Kwitt R, Vialard FX (2019) Metric learning for image registration. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp 8455–8464. https://doi.org/10.1109/CVPR.2019.00866

Oliveira F, Tavares J (2014) Medical image registration: A review. Computer methods in biomechanics and biomedical engineering 17:73–93. https://doi.org/10.1080/10255842.2012.670855

Olivo JC, Deubler J, Boulin C (1995) Automatic registration of images by a wavelet-based multiresolution approach. Proceedings of SPIE - The International Society for Optical Engineering 2569. https://doi.org/10.1117/12.217579

Paquin D, Levy D, Schreibmann E, et al (2006) Multiscale image registration. Mathematical biosciences and engineering : MBE 3:389–418. https://doi.org/10.3934/mbe.2006.3.389

Paquin D, Levy D, Xing L (2007) Hybrid multiscale landmark and deformable image registration. Mathematical biosciences and engineering : MBE 4:711–37. https://doi.org/10.3934/mbe.2007.4.711

Paquin D, Levy D, Xing L (2008) Multiscale deformable registration of noisy medical images. Mathematical biosciences and engineering : MBE 5:125–44. https://doi.org/10.3934/mbe.2008.5.125

Risser L, Vialard FX, Wolz R, et al (2011) Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping. IEEE transactions on medical imaging 30:1746–59. https://doi.org/10.1109/TMI.2011.2146787

Rueckert D, Sonoda L, Hayes C, et al (1999) Nonrigid registration using free-form deformations: application to breast mr images. IEEE Transactions on Medical Imaging 18(8):712–721. https://doi.org/10.1109/42.796284

Shen Z, Vialard FX, Niethammer M (2019) Region-specific diffeomorphic metric mapping. In: Wallach H, Larochelle H, Beygelzimer A, et al (eds) Advances in Neural Information Processing Systems, vol 32. Curran Associates, Inc.

Sommer S, Lauze F, Nielsen M, et al (2012) Sparse multi-scale diffeomorphic registration: The kernel bundle framework. J of Mathematical Imaging and Vision 46. https://doi.org/10.1007/s10851-012-0409-0

Staring M, Heide U, Klein S, et al (2009) Registration of cervical mri using multifeature mutual information. IEEE transactions on medical imaging 28:1412–21. https://doi.org/10.1109/TMI.2009.2016560

Sun W, Niessen WJ, Klein S (2014) Wavelet based free-form deformations for nonrigid registration. In: Ourselin S, Styner MA (eds) Medical Imaging 2014: Image Processing, International Society for Optics and Photonics, vol 9034. SPIE, pp 934–939, https://doi.org/10.1117/12.2043846

Tadmor E, Nezzar S (2004) A multiscale image representation using hierarchical (bv,l^2) decompositions. Multiscale Model Simul 2. https://doi.org/10.1137/030600448

Tan M, Qiu A (2016) Large deformation multiresolution diffeomorphic metric mapping for multiresolution cortical surfaces: A coarse-to-fine approach. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society 25. https://doi.org/10.1109/TIP.2016.2574982

Thompson DW (1992) On Growth and Form. Canto, Cambridge University Press, https://doi.org/10.1017/CBO9781107325852

Torbati N, Ayatollahi A (2017) A new method for non-rigid registration of mri images. In: 2017 Iranian Conference on Electrical Engineering (ICEE), pp 91–94, https://doi.org/10.1109/IranianCEE.2017.7985186

Trouvé A (1998) Diffeomorphisms groups and pattern matching in image analysis. Int J Comput Vision 28(3):213–221. https://doi.org/10.1023/A:1008001603737

Wu YT, Kanade T, Li C, et al (2000) Image registration using wavelet-based motion model. International Journal of Computer Vision 38:129–152. https://doi.org/10.1023/A:1008101718719