Non-rigid registration of magnetic resonance imaging of brain.

Clément Chesseboeuf[12], Hermine Biermé[1], Julien Dambrine [1], Carole Guillevin[2], Rémy Guillevin[2].
[1] LMA Université de Poitiers,
[2] DACTIM, CHU de Poitiers.
e-mail: clement.chesseboeuf@math.univ-poitiers.fr

***Abstract*— This document present a non-rigid registration framework for the use of brain magnetic resonance (MR) images comparison. More precisely we want to compare pre-operative and post-operative MR images in order to assess the deformation due to the surgical removal of tumor. Consequently, we propose an application of the theory developed in [3] associated with a new matching criterion based on a gradient representation in the dual of a RKHS (Reproducing Kernel Hilbert Space). Moreover, all objects are defined from a periodic point of view, allowing the construction of an efficient algorithm. Numerical results are presented**

***Keywords*— Image registration, greedy algorithm , matching criterion , diffeomorphism, gradient flow.**

I. Registration problem

Let u_0 and u_T a couple of "source" and "target" images. For instance, a couple of pre/post-operative MR images:

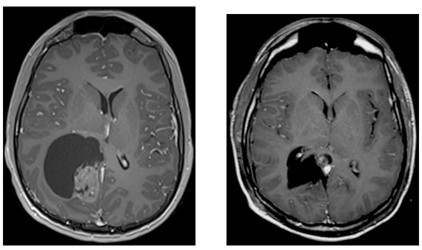

Fig. 1. Left: source image u_0 . Right: target image u_T.

Registration problem aims at construct a transformation ϕ, acting on the ambient space, such that:

$$u_0 \circ \phi^{-1} = u_T. \tag{1}$$

Since an exact matching is impossible in general, we will instead try to find ϕ such that:

$$u_0 \circ \phi^{-1} \approx u_T.$$

This proximity has to be quantified by a discrepancy measurement $d(\cdot, \cdot)$ that needs to be defined. This leads to the following optimisation problem:

$$\phi^* = \underset{\phi \in \mathcal{A}}{\arg\min}\ d(u_0 \circ \phi^{-1}, u_T) \tag{2}$$

In order to define this problem mathematically we need:

1) A matching criterion $J(\phi) = d(u_0 \circ \phi^{-1}, u_T)$, to quantify the discrepancy between the target image and the deformed source image.
2) A set of deformations $\mathcal{A}$ suitable for optimization.
3) An efficient optimisation process.

Image registration can be performed in various ways, an exhaustive description can be found in [9]. Note that the classification of registration methods is generally based on the three points mentioned above. In this paper, deformations are modelled by a group of diffeomorphisms (elastic transformations) on which we performe gradient steps in order to reach the minimum of a matching criterion. This model know as the "Greedy algorithm" was firstly described in [2]. A theorical explanation was developed in [3]. This is a non-parametric method able to deform the whole domain without using a set of control points (as with Radial basis functions (RBF) model).

Most of registration models can be embedded in a variational problem, such as the following:

$$u^* = \underset{u \in H}{\arg\min}\ d(\, u_0 \circ (Id + u)^{-1}\,,\, u_T\,) \,+\, R(u). \tag{3}$$

The optimisation variable u is a vector field:

$$u : \mathbb{R}^2 \longmapsto \mathbb{R}^2$$

belonging to an Hilbert space. The main difficulty is to obtain a solution $u^\star$ leading to a smooth and one-to-one transformation $Id + u^\star$. Therefore, the additonal term $R(u)$ is introduced to correct the ill-posedness of the problem. This term includes a regularisation part and a constraint on the differential $D(Id + u)$ in order to recover smooth and one-to-one transformation. Positive results about this formulation can be found in [4] and [5].

The originality of the model we will use is that it doesn't need (at first sight) any additional term. Indeed, the optimisation variable is constructed directly as a smooth diffeomorphism. This is a great advantage to performe optimisation in a set composed exclusively of elastic transformations.

The contribution of this paper is a derivation of this model in a periodic framework. Moreover, we present and use a new matching criterion constructed in the idea of current measures.

In section III we describe our matching criterion. Section IV takes a brief look at deformation space and section V is focused on the optimisation process.

A last requirement is that each object we define is suitable for numerical computation. For this purpose we introduce the reproducing kernel Hilbert spaces (RKHS) in the following section.

II. Reproducing Kernel Hibert Spaces (RKHS).

This section aims at describing RKHS properties. There are lots of reason for using this space as a modeling building block. On one hand, many of well known Hilbert spaces are RKHS; on the other hand, these spaces are convenient from numerical point of view. There is a large literature about RKHS and more details can be found in [11].

Definition 1. Let X be a set, for example the 2-dimensional torus:

$$\mathbb{T}^2 = (\mathbb{R}^2/\mathbb{Z}^2)$$

and $(H\,,\langle\cdot\,|\,\cdot\rangle_H)$ an Hilbert space of real functions defined on X. For each $x \in X$, there is an algebraic linear form that evaluates function at x:

$$\delta_x : f \to f(x) \quad \forall f \in H.$$

H is a reproducing kernel Hilbert space if:

$$\delta_x \in H^{'} \quad \forall x \in X.$$

More precisely, for all $x \in X$, there is a constant $M_x > 0$ such that:

$$|f(x)| \leqslant M_x \|f\|_H, \quad \forall f \in H.$$

The Riesz representation theorem implies that for all $x \in X$, there is an unique function $K_x \in H$ with the reproducing property:

$$f(x) = \delta_x(f) = \langle f\,|\,K_x\rangle_H\,, \quad \forall f \in H.$$

Then, exchanging f and K_y we define a function:

$$K : X \times X \to \mathbb{R}$$

by

$$K(x,y) = \langle K_x\,|\,K_y\rangle_H = K_x(y), \quad \forall\,(x,y) \in X^2.$$

This is the reproducing kernel of H.

In the rest of this paper, we will use RKHS as a test space. In other words, a function (corresponding to a brain MRI) will be represented by a element of the dual H' of H. In the RKHS framework, a good knowledge of the kernel function allows many explicit calculations. As an example, consider the linear isomorphism described by Riesz theorem:

$$R_H = H^{'} \to H.$$

Given any linear form $\varphi \in H^{'}$, there is a unique element $R_H(\varphi) \in H$ such that:

$$\langle R_H(\varphi)\,|\,f\rangle_H = \varphi(f), \quad \forall f \in H.$$

Moreover the Cauchy-Schwartz inequality yields:

$$\|\varphi\|^2_{H'} = \langle R_H(\varphi)\,|\,R_H(\varphi)\rangle_H\,.$$

In many cases, $R_H(\varphi)$ appears as the solution of a concrete problem. In the RKHS framework, if we know the kernel and the scalar product, $R_H(\varphi)$ is entirely recovered by the formula

$$R_H(\varphi)(x) = \delta_x\left(R_H(\varphi)\right) = \langle R_H(\varphi)\,|\,K_x\rangle_H\,.$$

III. Matching criterion.

A. *Current norm on curves.*

In [1], the author gives a method to put a norm on curves by representing them as a "current".

Let $\mathcal{C}$, be a curve in $\mathbb{R}^2$, and W an Hilbert space of vector fields. By testing the circulation of each $w \in W$ along $\mathcal{C}$, we obtain a linear form on W :

$$\begin{aligned} \gamma_c : W &\longrightarrow \mathbb{R} \\ w &\longmapsto \textstyle\int_{\mathcal{C}} w \cdot \overrightarrow{dl}\,, \end{aligned} \tag{4}$$

where $\overrightarrow{dl}$ is tangential line element of the curve $\mathcal{C}$.

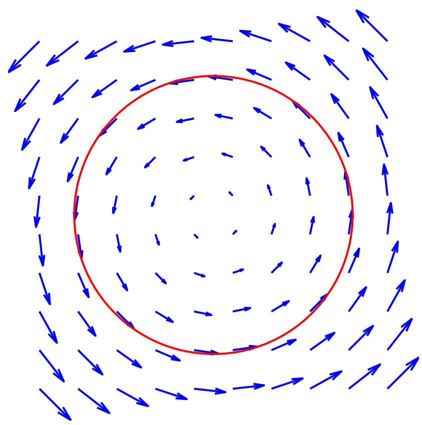

Fig. 2. An example a vector field inducing high circulation on a circle.

The mapping $\mathcal{C} \to \gamma_c$ yields a useful representation of curves, as an element of the dual vector space W'. Such an object is naturally normed by:

$$\|\gamma_c\|_{W'} = \sup_{\|w\|_W \leqslant 1} |\gamma_c(w)|$$

Let W be the cartesian product of two identical RKHS with a Kernel K_W and $c(t)$ be a parametrisation of $\mathcal{C}$, we compute this norm by using the formula

$$\|\gamma\|^2_{W'} = \int_0^T \int_0^T \langle c'(s)\,|\,c'(t)\rangle\, K_W(c(s),c(t))\;ds\,dt.$$

Note that the determination of a parametrisation $c(t)$ from imaging requires a curve extraction step which is particularly challenging in the case of brain images.

B. *Extended version of the current norm.*

The approach described above is very helpful for a geometrical representation of images. We want to get rid of the curve extraction step and work directly on greyscale images as functions. In the following section we show that it is possible, through the Stokes theorem, to compute a current norm based on curves without the geometrical extraction step.

For the sake of simplicity we first consider binary images. Let $u = \mathbb{1}_D$ be the characteristic function of a bounded regular domain $D \subset \mathbb{T}^2$, with the notation $\partial D = \mathcal{C}$:

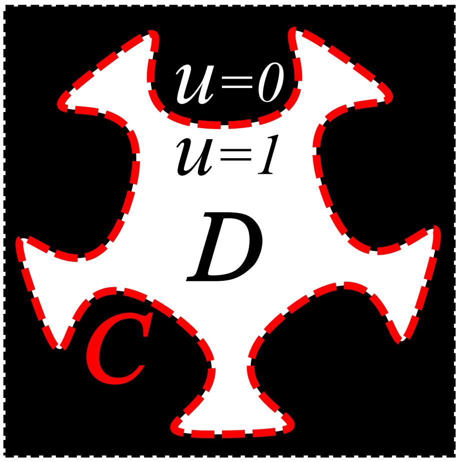

Fig. 3. Example of a binary image representing a domain D.

Let $v : \mathbb{T}^2 \to \mathbb{R}^2$ be a periodic vector field. The curl of v is definied by:

$$(\nabla \times v)\,(x) = \partial_1 v_2(x) - \partial_2 v_1(x).$$

The following theorem is verified under regularity hypothesis on D.

Theorem 1 (Stokes)**.**

$$\int_{\mathcal{C}} w\ \vec{dl} = \int_D \nabla \times w = \int_{\mathbb{T}^2} (\nabla \times w)\,(x)\,u(x)\ dx. \qquad (5)$$

From (5) we remark that the computation of any current can be performed with an integral over the whole domain. We will now generalise this notion for any image with greyscale values.

Let $u : \mathbb{T}^2 \to \mathbb{R}^+$ be a generic non negative function. To make the following explanation work, we assume that u is regular and compactly supported.

Each level $r \in \mathbb{R}^+$ yields a domain $D(r) = \{x,\ u(x) \geqslant r\}$ and almost each set $D(r)$ yields a curve $\partial D(r)$ (assertion based on Sard's theorem, see [10]). Consequently we have a one parameter family of curves and the dual representation described by (4) leads to the one parameter family of linear forms:

$$\begin{aligned} \mathbb{R}^+ &\longrightarrow W' \\ r &\longmapsto \gamma_r . \end{aligned}$$

such that:

$$\gamma_r(w) = \int_{\partial D(r)} w\ \vec{dl_r}.$$

By integrating this family we obtain an other linear form defined by:

$$\begin{aligned} \Gamma_u(w) &= \int_{\mathbb{R}^+} \gamma_r(w)\ dr \\ &= \int_{\mathbb{R}^+} \int_{\mathbb{T}^2} (\nabla \times w)\,(x)\,\mathbb{1}_{\{u \geqslant r\}}(x)\ \ dx\,dr\,. \end{aligned} \qquad (6)$$

$\{u \geqslant r\}$ is bounded because u is compactly supported. Then, if $\nabla \times w$ is continuous (or at least locally integrable) the above integrale is well defined an we can use a Fubini inversion to obtain:

$$\Gamma_u(w) = \int_{\mathbb{T}^2} (\nabla \times w)\,(x)\,u(x)\ dx. \qquad (7)$$

Finally, Γ_u is a representation of the function u. We will use the notation:

$$\|u\|_{cW} = \|\Gamma_u\|_{W'}.$$

This norm is a measure of the gradient (actually of a rotated gradient). Therefore, global contrast deviation (or addition of a constant) is not take into account. Moreover, there is a regularisation phenomenon, due to the particular scalar product defined on W. Regularity of the matching criteria is a determinant characteristic to avoid local minima in steepest descent algorithm.

Let us give a more precise description of the model. Brain images can be extended as periodic functions, since they are compactly supported. As a consequence, we derive a periodic model based on a RKHS W, composed of periodic vector fields. To this aim we use periodic Sobolev spaces H^s_{per}:

$$W = H^s_{per} \times H^s_{per}\ ,$$

where, $s > 1$ is a positive parameter controlling the regularity of $w \in W$.

Let $k \in \mathbb{Z}^2$ and $f : \mathbb{T}^2 \to \mathbb{C}$. Then, $c_k(f)$ denotes the "k-th" Fourier coefficient of f.

W is an Hilbert space for the scalar product:

$$\langle w \,|\, v \rangle_W = \sum_{k \in \mathbb{Z}^2} (1 + |k|^2)^s \left(c_k(w_1)\overline{c_k(v_1)} + c_k(w_2)\overline{c_k(v_2)} \right)$$

We know, by Sobolev spaces theory that a parameter $s > 1$ leads to continuous function in H^s_{per}. Therefore, by fixing an $s > 1$, each coordinate belongs to an RKHS. The Kernel of this RKHS is:

$$K_W(x, y) = \sum_{k \in \mathbb{Z}^2} \frac{1}{(1 + |k|^2)^s}\ e^{2i\pi k \cdot (x - y)}. \qquad (8)$$

We notice that this kernel is stationary, hence there is a function f such that:

$$K_W(x, y) = f(x - y)\,.$$

Finally, by using RKHS properties (see the calculations in appendix) with the classical notation:

$$\Delta f(x) = \frac{\partial^2 f}{\partial x_1^2}(x) + \frac{\partial^2 f}{\partial x_2^2}(x),$$

we obtain an explicit formula for the computation of the current norm:

$$\|\Gamma_u\|_{W'}^2 = -\left\langle u \,|\, \Delta f \star u \right\rangle_{L^2} . \tag{9}$$

This expression can be quickly computed by using FFT. Moreover, when u represents a binary image describing the contour $\mathcal{C}$, we have:

$$\|\Gamma_u\|_{W'} = \|\gamma_c\|_{W'}$$

IV. Deformation space.

The fundamental hypothesis is that our deformation ϕ defined in (1) is a diffeomorphism, that is to say, a smooth and one-to-one with a smooth inverse map. This concept includes a large class of transformations, from very simple mappings such as translations to much more complex transformations (large non-linear deformations).

The main difficulty here is that there is no parametrisation of a diffeomorphism set which would be appropriate for optimization. For instance there is no guarantee that the sum of two diffeomorphisms is a diffeomorphism. Hence our set is not a vector space and we dont have any natural notion of direction (for example for gradient descent methods). However this set of deformations has a group structure analogous to a lie group whith its associated lie algebra (here, a set of smooth vector fields). In the rest of this section, we show how to exploit this structure.

Let $(v_t)_{t\in[0,\epsilon]}$ be a time dependent vector field, we know, from Cauchy-Lipschitz theory that v_t yields a time dependent diffeomorphism ϕ_t, solution of the ODE:

$$\begin{cases} \partial_t \phi_t(x) = v_t(\phi_t(x)) & t \in [0, \epsilon] \\ \phi_0(x) = x \quad \forall x \in \mathbb{T}^2. \end{cases} \tag{10}$$

Let V be another Hilbert space of vector fields. By using $v_t \in V$ as a parameter, we can construct a diffeomorphism group $\mathcal{A}_V$ suitable for optimisation. This is the theory develop in [3] in which the author, after a rigorous construction, gives two methods for solving matching-type problems.

The second method is based on the possibility to define the gradient flow of a function on the group $\mathcal{A}_V$. The following section describes an application using the above matching criterion.

V. Gradient flow on diffeomorphism group.

A. *An instructive example with translation group*

Let u_T be a translation of u_0 by $b \in \mathbb{T}^2$:

$$u_T(x) = \tau_b u_0(x) = u_0(x - b).$$

For each $a \in \mathbb{T}^2$, our matching criterion gives a score:

$$J(a) = \frac{1}{2}\|\tau_a u_0 - u_T\|_{W'}^2,$$

where J is a positive function and b is a global minimum of J. In order to find b, we will construct a path $t \to a(t)$ on which $J(a(t))$ is decreasing. A good way for achieving this target is to solve the following ODE, which is a continuous analogue of the discrete gradient descent algorithm.

$$\begin{cases} a'(t) = -\nabla_{a(t)} J & t \in \mathbb{R}^+ \\ a(0) = 0. \end{cases} \tag{11}$$

From (9), we have:

$$J(a) = -\frac{1}{2}\left\langle \tau_a u_0 - u_T \,|\, \Delta f \star (\tau_a u_0 - u_T) \right\rangle_{L^2} . \tag{12}$$

Under regularity assumptions, we obtain the following characterisation of the gradient :

$$\nabla_a J = \left\langle \nabla \tau_a u_0 \,|\, \Delta f \star (\tau_a u_0 - u_T) \right\rangle_{L^2} .$$

Let $a(t)$ be a solution of (11), then we have

$$\frac{d}{dt} J(a_t) = -\|\nabla_{a(t)} J\|^2 \leqslant 0.$$

The following figures show some gradient steps on an artificial example of a translated image.

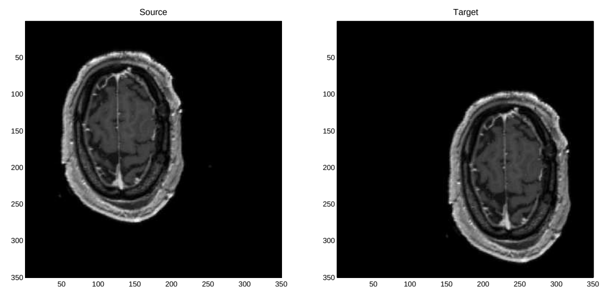

Fig. 4. Source and target image.

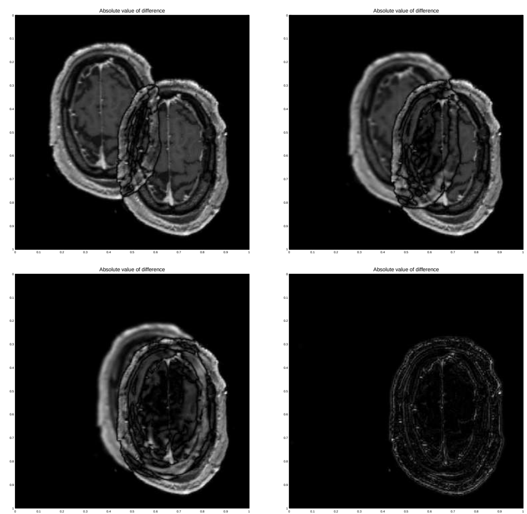

Fig. 5. From left to right and from top to bottom, gradient flow evolution of $|\tau_{a(t)} u_0 - u_T|$.

Consider translations as a Lie group, the success of previous method is due to the Riemannian structure of this group. More precisely it is due to the possibility of defining a gradient. Now, we will extend this procedure to a more general set of diffeomorphisms. In this case, the notion of tangent space and gradient has to be defined.

B. Gradient flow on an infinite dimensional diffeomorphism group.

In this section, we want to reproduce the previous procedure by replacing the translation variable a by a general diffeomorphism ϕ. In this case the matching criterion is:

$$J(\phi) = \frac{1}{2}\|u_0 \circ \phi^{-1} \ - \ u_T\|^2_{W'}.$$

In order to achieve a good conversion from a to ϕ, we have to use a little more "Riemannian" description.

Given a state ϕ we are looking for a direction in which J is decreasing. Let ϕ_t be a time dependent diffeomorphism such that $\phi_0 = \phi$.

There is no problem to interpret this object as a curve passes through the "point" ϕ, at time $t = 0$. The notion of direction or "tangent vector" is mush more ambiguous but can be defined by using an ODE

$$\begin{cases} \partial_t \phi_t(x) = v_t(\phi_t(x)) & t \in [0, \epsilon] \\ \phi_0(x) = \phi(x) & \forall x \in \mathbb{T}^2. \end{cases} \tag{13}$$

where v_t is a one parameter family of vector fields belongs to an Hilbert space. Formally,

$$v_0 \circ \phi(x) = \frac{d}{dt}\phi_t(x)_{|t=0}$$

as to be considered as a tangent vector to ϕ. Consequently, we can compute the directional derivative of J in this direction:

$$\begin{gathered} \tfrac{d}{dt}J(\phi_t)_{|t=0} \\ = \\ \left\langle v_0 \ \middle| \ \nabla\left(u_0 \circ \phi^{-1}\right) \times \Delta f \star (u_0 \circ \phi^{-1} - u_T)\right\rangle_{L^2}, \end{gathered}$$

Note that we follow the same procedure as those for translation group.

Here, we have to be careful not to interpret

$$\nabla\left(u_0 \circ \phi^{-1}\right) \times \Delta f \star (u_0 \circ \phi^{-1} - u_T)$$

as the gradient of J. It could be the case if we are working with the Hilbert space $V = L^2(\mathbb{T}^2, \mathbb{R}^2)$. However we need more regularity to give sense to system (13). Admissibility conditions are described by Cauchy-Lipschitz theorem (local lipschitz vector field) and are completed if we have some continous embedding in a Banach space of regular functions.

$$(V, \|\cdot\|_V) \quad \subset \quad \left(\mathcal{C}^p(\mathbb{T}^2, \mathbb{R}^2), \|\cdot\|_{p,\infty}\right).$$

with $p \geqslant 1$.

Recall that v_0 is an element of an Hilbert space V, and assume that the mapping

$$\begin{aligned} &V \longrightarrow \mathbb{R} \\ &v_0 \longmapsto \left\langle v_0 \ \middle| \ \nabla\left(u_0 \circ \phi^{-1}\right) \times \Delta f \star (u_0 \circ \phi^{-1} - u_T)\right\rangle_{L^2} \end{aligned}$$

is a continous linear form on V. With an other use of Riesz theorem, there is a unique vector field denoted $\tilde{v}$ such that:

$$\begin{gathered} \langle v_0 \,|\, \tilde{v}\rangle_V \\ = \\ \left\langle v_0 \ \middle| \ \nabla\left(u_0 \circ \phi^{-1}\right) \times \Delta f \star (u_0 \circ \phi^{-1} - u_T))\right\rangle_{L^2}. \end{gathered}$$

Thus, if we choose the direction $v_0 = -\tilde{v} \in V$, we obtain

$$\frac{d}{dt}J(\phi_t)_{|t=0} = -\|\tilde{v}\|^2_V \leqslant 0.$$

Defining V as an other RKHS, we can easily compute $\tilde{v}$ by using an explicit linear operator R_V based on convolution with the reproducing kernel of V:

$$\tilde{v} = R_V\left(\nabla\left(u_0 \circ \phi^{-1}\right) \times \Delta f \star (u_0 \circ \phi^{-1} - u_T)\right).$$

Finally we introduce

$$\nabla^V_\phi J = R_V\left(\nabla\left(u_0 \circ \phi^{-1}\right) \times \Delta f \star (u_0 \circ \phi^{-1} - u_T)\right) \circ \phi.$$

And the gradient flow equation is

$$\begin{cases} \partial_t \phi_t(x) = -\nabla^V_{\phi_t} J\,(x) \\ \phi_0(x) = x. \end{cases} \tag{14}$$

The following figure show an example of gradient flow progression. Source and target images are characteristic functions of a tumorous area, before and after surgical removal.

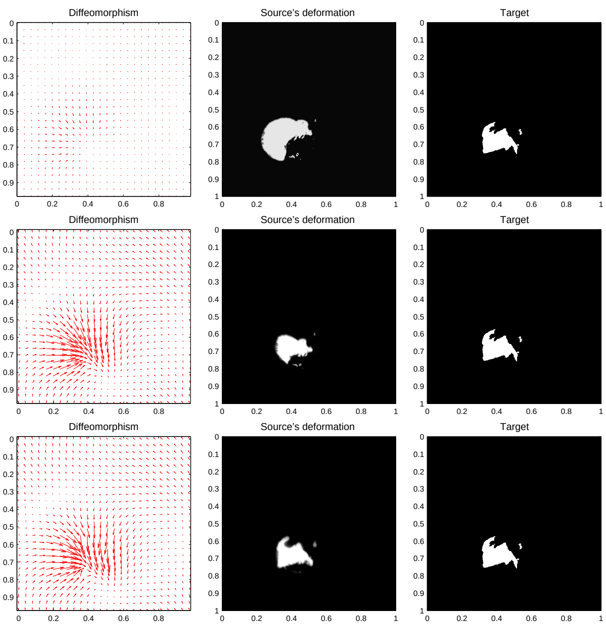

Fig. 6. Gradient flow evolution.
First column : ϕ_t improvement. Second column : $u_0 \circ \phi_t^{-1}$. Third : u_T.

VI. Conclusion

We have described a registration algorithm that takes advantage of the Riemannian structure of the group of diffeomorphisms. This approach, sometimes referred as "greedy", leads to an efficient registration process especially in the periodic framework use in this paper. Indeed, for $n \times m$ images, each gradient step has a computational cost of $\mathcal{O}(n\times m)log(n\times m)$ operations. As a result, in the case of 512×512 images such as those presented in figure (6), the computation of a solution takes a few minutes with a Matlab codes. The use of the matching criterion introduced in this paper doesn't increase the computational cost (as illustrated by formula (9)). Futhermore, this criterion seems better suited for matching than the classical L^2 norm. On one hand, it is based on image gradient, thus focused on structure (contours, critical points...). On the other hand, the regularisation helps prevent local minima from early stopping the algorithm.

VII. Appendix

In this appendix we detail the computation of the current norm (9) based on a greyscale function u. Recall the expression of the current measure Γ_u:

$$\Gamma_u(w) = \int_{\mathbb{T}^2} \nabla \times w(x)\, u(x)\, \mathrm{d}x\,. \quad (15)$$

By denoting $w = (w_1, w_2)$ and $\nabla = (\partial_1, \partial_2)$, we have:

$$\Gamma_u(w) = \int_{\mathbb{T}^2} [\partial_1 w_2(x) - \partial_2 w_1(x)]\, u(x)\, \mathrm{d}x\,. \quad (16)$$

Since w_1 and w_2 both belong to the same RKHS H with the kernel K_W, we use the reproducing property to obtain:

$$\Gamma_u(w) = \int_{\mathbb{T}^2} [\partial_1 \langle K_W(x,\cdot) \,|\, w_2 \rangle_H - \partial_2 \langle K_W(x,\cdot) \,|\, w_1 \rangle_H]\, u(x)\, \mathrm{d}x\,. \quad (17)$$

Using Fubini's theorem, we write $\Gamma_u(w)$ as a dot product in H:

$$\Gamma_u(w) = \left\langle \int_{\mathbb{T}^2} u(x)\, \partial_1 K_W(x,\cdot)\mathrm{d}x \,|\, w_2 \right\rangle_H - \left\langle \int_{\mathbb{T}^2} u(x)\, \partial_2 K_W(x,\cdot)\mathrm{d}x \,|\, w_1 \right\rangle_H, \quad (18)$$

which can be rewritten as a dot product in W:

$$\Gamma_u(w) = \left\langle \int_{\mathbb{T}^2} u(x)\, \nabla^\perp K_W(x,\cdot)\mathrm{d}x \,|\, w \right\rangle_W, \quad (19)$$

where $\nabla^\perp = (-\partial_2, \partial_1)$ is the "turned" gradient operator. This defines a continuous linear form on W. Thanks to the Cauchy-Schwartz inequality, the norm of Γ_u is given by:

$$\|\Gamma_u\|_{W'} = \| \int_{\mathbb{T}^2} u(x)\, \nabla^\perp K_W(x,\cdot)\, \mathrm{d}x \,\|_W\,, \quad (20)$$

which leads to:

$$\|\Gamma_u\|^2_{W'} = \left\langle \int_{\mathbb{T}^2} u(x)\, \nabla^\perp_x K_W(x,\cdot)\, \mathrm{d}x \,|\, \int_{\mathbb{T}^2} u(y)\, \nabla^\perp_y K_W(y,\cdot)\, \mathrm{d}y \right\rangle_W. \quad (21)$$

where the operator $\nabla^\perp_x$ (respectively $\nabla^\perp_y$) is the turned gradient with respect to x (respectively y). Again, using Fubini's theorem, we obtain:

$$\|\Gamma_u\|^2_{W'} = \int_{\mathbb{T}^2}\int_{\mathbb{T}^2} u(x)u(y)\, \nabla^\perp_x \cdot \nabla^\perp_y \langle K_W(x,\cdot) \,|\, K_W(y,\cdot) \rangle_H\, \mathrm{d}y\mathrm{d}x \quad (22)$$

The reproducing kernel property allows us to simplify this expression to:

$$\|\Gamma_u\|^2_{W'} = \int_{\mathbb{T}^2}\int_{\mathbb{T}^2} u(x)u(y)\, \nabla^\perp_x \cdot \nabla^\perp_y K_W(x,y)\, \mathrm{d}y\mathrm{d}x \quad (23)$$

Since the kernel is supposed to be stationary (*i.e.* $K_W(x,y) = f(x-y)$), we have:

$$\nabla^\perp_x \cdot \nabla^\perp_y K_W(x,y) = -\Delta f(x-y)\,, \quad (24)$$

hence:

$$\|\Gamma_u\|^2_{W'} = -\int_{\mathbb{T}^2}\int_{\mathbb{T}^2} u(x)u(y)\, \Delta f(x-y)\, \mathrm{d}y\mathrm{d}x\,. \quad (25)$$

Finally we obtain the expression (9) by noticing the convolution product in the above expression.

References

[1] J.Glaunès. *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique.* PhD thesis, Universite Paris 13, 2005.

[2] Christensen, G.E, Rabbitt,R.D,and Miller, M.I. 1996. *Deformable templates using large deformation kinematics* IEEE Transactions on Image Processing, 5(10):1435-1477.

[3] A.Trouvé. *An infinite dimensional group approach for physics based models in patterns recognition*, 1995.

[4] F.J.P Richard. *Résolution de problèmes hyperélastiques de recalage d'images.* C.R Acad Sci. Paris, Ser. I 335 (2002) 295-299.

[5] M. Droske and M. Rumpf. *A variational approach to non-rigid morphological registration.* SIAM Appl. Math., 64(2):668–687, 2004.

[6] I.Miller, A.Trouvé, and L.Younes. *On the metrics and Euler-Lagrange equations of computational anatomy. Ann. Review of Biomedical Engineering*, 4 (2002), pp. 375–405.

[7] I. Miller, A.Trouvé, and L.Younes. *Geodesic Shooting for Computational Anatomy.* Journal of Mathematical Imaging and Vision, vol. 24, no. 2, pp. 209–228, 2006.

[8] L.Younes. *Shapes and Diffeomorphisms.* Springer, 2010.

[9] A.Sotiras, C.Davatzikos, N.Paragios. *Deformable Medical Image Registration: A Survey* IEEE Transaction on Medical Imaging, vol 32, no 7, July 2013.

[10] J.Milnor. *Topology from a Differentiable Viewpoint.* Princeton University Press, 1997.

[11] N.Aronszajn. *Theory of reproducing kernels.* Trans. Amer. Math. Soc. 63:337–404, 1950.

[12] H.Brezis. *Analyse fonctionnelle.*