# Generalized Gradients: Priors on Minimization Flows

G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, O. Faugeras
Odyssee Lab* ENS/INRIA/ENPC
Paris/Sophia-Antipolis/Champs-sur-Marne, France

## Abstract

*This paper tackles an important aspect of the variational problem underlying active contours: optimization by gradient flows. Classically, the definition of a gradient depends directly on the choice of an inner product structure. This consideration is largely absent from the active contours literature. Most authors, explicitly or implicitely, assume that the space of admissible deformations is ruled by the canonical $L^2$ inner product. The classical gradient flows reported in the literature are relative to this particular choice. Here, we investigate the relevance of using (i) other inner products, yielding other gradient descents, and (ii) other minimizing flows not deriving from any inner product. In particular, we show how to induce different degrees of spatial consistency into the minimizing flow, in order to decrease the probability of getting trapped into irrelevant local minima. We report numerical experiments indicating that the sensitivity of the active contours method to initial conditions, which seriously limits its applicability and efficiency, is alleviated by our application-specific spatially coherent minimizing flows. We show that the choice of the inner product can be seen as a prior on the deformation fields and we present an extension of the definition of the gradient toward more general priors.*

## 1. Introduction

Many problems in computer vision can advantageously be cast in a variational form, i.e. as a minimization of an energy functional. In this paper, we focus on variational methods dedicated to the recovery of contours. In this case, the problem amounts to finding a contour which corresponds to a global minimum of the energy. Unfortunately, in most cases, the exact minimization of the energy functional is computationally unfeasible due to the huge number of unknowns.

The *graph cuts* method is a powerful energy minimization method which allows to find a global minimum or a strong local minimum of an energy. In the last few years, this method has been successfully applied to several problems in computer vision, including stereovision [17] and image segmentation [5]. However, it has a severe limitation: it cannot be applied to an arbitrary energy function [18], and, when applicable, is computationally expensive.

Hence, in most cases, a suboptimal strategy must be adopted. A common minimization procedure consists in evolving an initial contour, positioned by the user, in the direction of steepest descent of the energy. This approach, known in the literature as *active contours* or *deformable models*, was pioneered by Kass. et

---

*Guillaume.Charpiat@di.ens.fr, Pierre.Maurel@di.ens.fr, Jean-Philippe.Pons@sophia.inria.fr, Renaud.Keriven@certis.enpc.fr, Olivier.Faugeras@sophia.inria.fr

al. in [16] for the purpose of image segmentation. Since, it has been applied in many domains of computer vision and image analysis (image segmentation [6], surface reconstruction [35, 11], stereo reconstruction [12, 15, 13], etc.).

However, due to the highly non-convex nature of most energy functionals, a gradient descent flow is very likely to be trapped in a local minimum. Also, this local minimum depends on the position of the initial contour. If the latter is far from the expected final configuration, the evolution may be trapped in a completely irrelevant state. This sensitivity to initial conditions seriously limits the applicability and efficiency of the active contours method.

We detail in section 2 the general gradient descent process so as to emphasize the crucial role of the inner product. After an abstract study in section 3 on how to handle inner products and minimizing flows, we propose, in section 4, various inner products and show how they induce different degrees of spatial coherence in the minimizing flow with numerical examples of shape warping in section 5. In section 6, a rewriting of the usual definition of the gradient shows how the choice of an inner product can be seen as a way to introduce a prior on the deformation fields, and this leads us to a natural extension of the notion of gradient to more general priors.

## 2. Minimization and inner product

In the following we consider a shape $\Gamma$, seen as a manifold of dimension $k$ embedded in $\mathbb{R}^n$, for example a planar curve or a surface in the space $\mathbb{R}^3$. We denote by $E(\Gamma)$ the energy functional to be minimized. In order to define the gradient of the energy functional, the first step is to compute its Gâteaux derivatives $\delta E(\Gamma, v)$ in all directions, i.e. for all admissible velocity fields $v$ defined on the shape $\Gamma$ with values in $\mathbb{R}^n$. The deformation space, set of all these fields $v$, can be seen as the tangent space of $\Gamma$, considered itself as a point in the manifold of all admissible shapes.

$$\delta E(\Gamma, v) \stackrel{def}{=} \lim_{\epsilon \to 0} \frac{E(\Gamma + \epsilon v) - E(\Gamma)}{\epsilon} \ . \tag{1}$$

where $\Gamma + \epsilon v$ is the shape defined, for any parametrization of $\Gamma$, say $\mathcal{P}_\Gamma : \sigma \in S \subset \mathbb{R}^k \mapsto \mathcal{P}_\Gamma(\sigma) \in \Gamma \subset \mathbb{R}^n$, by the parametrization $\mathcal{P}_{\Gamma + \epsilon v} : \sigma \in S \mapsto \mathcal{P}_{\Gamma + \epsilon v}(\sigma) = \mathcal{P}_\Gamma(\sigma) + \epsilon v(\mathcal{P}_\Gamma(\sigma)) \in \mathbb{R}^n$.

Then, we would like to pick the gradient as the direction of steepest descent of the energy. However, it is not yet possible at this stage: to be able to assess the steepness of the energy, the deformation space needs additional structure, namely an inner product introducing the geometrical notions of angles and lengths. This consideration is largely absent from the active contours literature: most authors, explicitly or implicitely, assume that the deformation space is ruled by the canonical $L^2$ inner product on $\Gamma$, which is, for two deformation fields $u$ and $v$:

$$\langle u \,|\, v \rangle_{L^2} = \frac{1}{|\Gamma|} \int_\Gamma u(\mathbf{x}) \cdot v(\mathbf{x}) \, d\Gamma(\mathbf{x}) \ ,$$

where $d\Gamma(\mathbf{x})$ stands for the area element of the contour so that the integral over $\Gamma$ is intrinsic and does not depend on the parametrization.

Here, for sake of generality, we model the space of admissible deformations as an inner product space $(F, \langle \,|\, \rangle_F)$. If there exists a deformation field $u \in F$ such that

$$\forall v \in F, \ \delta E(\Gamma, v) = \langle u \,|\, v \rangle_F \ ,$$

then $u$ is unique, we call it the gradient of $E$ *relative to the inner product* $\langle \,|\, \rangle_F$, and we denote by $u = \nabla_F E(\Gamma)$. The existence of $u$ is related to the smoothness of $E$, or more exactly to the continuity of $\delta E(\Gamma, v)$ with respect to $v$ (Riesz representation theorem, see [27] for more details).

Clearly, each choice of inner product yields its own gradient. This is often neglected and most authors improperly refer to *the* gradient of the energy. Thus, the classical gradient flows reported in the literature (mean curvature flow, geodesic active contours [6, 14, 29], multi-view 3D reconstruction [12, 15, 13]) are relative to the $L^2$ inner product.

The gradient descent method consists in deforming an initial contour $\Gamma_0$ in the opposite direction of the gradient.

$$\begin{cases} \Gamma(0) = \Gamma_0 \\ \dfrac{d\Gamma}{dt} = -\nabla_F E(\Gamma) \end{cases} \tag{2}$$

The problem of the existence and the uniqueness of this minimizing flow is out of the scope of this article. Indeed, it is highly dependent on the properties of each particular energy functional. If this evolution exists, it decreases the energy:

$$\frac{dE(\Gamma)}{dt} = -\left\| \nabla_F E(\Gamma) \right\|_F^2 \leq 0 \; .$$

The standard choice for $F$ is the Hilbert space of square integrable velocity fields $L^2(\Gamma, \mathbb{R}^n)$ equipped with its canonical inner product. Very few authors in the active contours area have considered using other inner products, whereas this is an established technique in image registration [31]. Very recently, in the context of shape representation and analysis, [19, 32] have shown that slightly modifying the $L^2$ inner product allows to build well-behaved metrics in the space of curves; the particular case of the $H^1$ inner product has been simultaneously and independently investigated by us [8] and by Sundaramoorthi et al. [30].

The variations on the gradient descent theme, as in [4], will still be applicable to the new gradients we propose, since these methods are in fact not specific to the particular $L^2$ gradient.

Minimizing flows not deriving from any inner product, that is to say evolutions that decrease the energy, without any gradient interpretation, have also been overlooked so far. Note that any evolution fulfilling the condition

$$\frac{dE(\Gamma)}{dt} = \left\langle \nabla_F E(\Gamma) \left| \frac{d\Gamma}{dt} \right\rangle_F \leq 0 \tag{3}$$

is a candidate to solve the minimization problem. This idea, proposed in [29], is applied by the same authors to the alignment of curve in images in [23]: a complicated term in the gradient is safely neglected after checking that the evolution still decreases the energy.

The spirit of our work is different. We do not focus either on a specific inner product or on a particular energy functional. We rather explore general procedures to build some new inner products and to compute the associated gradients. We also address the design of non-gradient minimizing flows.

Our motivation is also different. Our primary concern in this work is the sensitivity of the active contours method to initial conditions. There are essentially two ways of dealing with this problem: positioning the initial contour very close to the expected final configuration, or using a multiresolution coarse-to-fine strategy, in other words running the optimization on a series of smoothed and subsampled contours and input data. In this paper, we pioneer a third way to tackle the problem of unwanted local minima: the careful design of the minimizing flow.

We do not modify the energy, hence the relief of the energy landscape and in particular the "number" of local minima remains unchanged. But by using an evolution that favors certain types of directions, we expect to decrease the probability of falling into unwanted energy basins.

Typically, in many applications, spatially coherent motions are to be preferred over erratic evolutions. For example, in the tracking problem, the object of interest is likely to have similar shapes in consecutive

frames. So if we initialize the contour with the result of the previous frame, it makes sense to encourage the motions which preserve its overall appearance. This way, it may be easier to dodge unexpected local low-energy configurations. A traditional $L^2$ gradient descent definitely does not have this desirable property since the $L^2$ inner product completely disregards the spatial coherence of the velocity field.

## 3. New Inner Products and New Flows

In this section, we suppose that the space $F$ of all admissible deformations of the shape $\Gamma$ is initially equipped with the inner product $\langle|\rangle_F$, for example in the standard case we would have $F = L^2$, and we study how to build new inner products or new minimizing flows from the given one.

### 3.1. Designing new inner products

**Definition 1.** *For any symmetric positive definite linear operator $\mathcal{L} : F \to F$, a new inner product can be defined by*

$$\langle u \,|\, v \rangle_{\mathcal{L}} = \langle \mathcal{L}u \,|\, v \rangle_F \ . \tag{4}$$

Here, for simplicity, we assume that the domain and the range of $\mathcal{L}$ are equal to $F$. A similar study is possible if they are strictly smaller than $F$, under certain conditions, using the Friedrichs extension of $\mathcal{L}$ (see [1] for details). But these technical details are out of the scope of this paper.

The following observation is central to our work:

**Proposition 2.** *If $\nabla_F E(\Gamma)$ exists and if $\mathcal{L}$ is also invertible, then $\nabla_{\mathcal{L}} E(\Gamma)$ also exists and we have*

$$\nabla_{\mathcal{L}} E(\Gamma) = \mathcal{L}^{-1} \left( \nabla_F E(\Gamma) \right) \ . \tag{5}$$

*Proof.* Indeed:

$$\begin{aligned}
\forall v \in F, \ \delta E(\Gamma, v) &= \langle \nabla_F E(\Gamma) \,|\, v \rangle_F \\
&= \langle \mathcal{L}\mathcal{L}^{-1} \nabla_F E(\Gamma) \,|\, v \rangle_F \\
&= \langle \mathcal{L}^{-1} \nabla_F E(\Gamma) \,|\, v \rangle_{\mathcal{L}} \ .
\end{aligned}$$

$\square$

The above procedure is of great practical interest because it allows to upgrade any existing $L^2$ gradient flow. However, it is not completely general in the sense than all inner products cannot be expressed in this form.

Nevertheless, if $F$ is a separable Hilbert space (i.e. complete with respect to the norm $\|\|_F$), the Riesz representation theorem tells us that any inner product $\langle|\rangle_{\mathcal{L}}$ such that

$$\exists C > 0, \ \forall u \in F, \ \|u\|_{\mathcal{L}} \leqslant C \|u\|_F$$

can be written in the form of equation (4). This suggests that our procedure accounts for a wide range of inner products.

## 3.2. Designing new minimizing flows

In this subsection, we follow the inverse approach. Instead of working with the inner product, we apply a linear operator $\mathcal{L} : F \to F$ to the gradient, and we study the properties of the resulting flow:

$$\frac{d\Gamma}{dt} = -\mathcal{L}\big(\nabla_F E(\Gamma)\big) . \tag{6}$$

This naturally sets up a hierarchy among different types of operators:

- if $\mathcal{L}$ is positive, the energy is non-increasing along the flow (6). Indeed,

$$\frac{dE(\Gamma)}{dt} = -\big\langle \nabla_F E(\Gamma) \,\big|\, \mathcal{L}\nabla_F E(\Gamma) \big\rangle_F \leqslant 0 .$$

- if $\mathcal{L}$ is positive definite, the energy strictly decreases along the flow (6) until a critical point of the original gradient flow (2) is reached.

- if $\mathcal{L}$ is symmetric positive definite and invertible, the flow (6) coincides with a gradient descent relative to the inner product $\langle | \rangle_{\mathcal{L}^{-1}}$, as defined in equation (4).

The third case is contained in Subsection 3.1. A useful example of the second case is given in Subsection 4.3.

## 3.3. Adding an orthogonal term

The rate of decrease of the energy when following the direction of descent $\frac{d\Gamma}{dt}$ is given by:

$$\frac{dE(\Gamma)}{dt} = \left\langle \nabla_F E(\Gamma) \,\middle|\, \frac{d\Gamma}{dt} \right\rangle_F \leqslant 0 .$$

In particular, for the usual evolution $\frac{d\Gamma}{dt} = -\nabla_F E(\Gamma)$, we have:

$$\frac{dE(\Gamma)}{dt} = -\|\nabla_F E(\Gamma)\|_F^2$$

If we denote by $v$ any vector field defined on $\Gamma$ such as $\langle \nabla_F E(\Gamma) \,|\, v \rangle_F = 0$, then adding such a vector field $v$ to the usual gradient descent term will not change the amount of decreased energy:

$$\frac{dE(\Gamma)}{dt} = \langle \nabla_F E(\Gamma) \,|\, -\nabla_F E(\Gamma) + v \rangle_F = -\|\nabla_F E(\Gamma)\|_F^2$$

so we can choose the field $v$ which we would like to add to the initial gradient. Rather than choosing $v = 0$ as usual, we could for example choose one, noted $\hat{v}$, that minimizes a regularizing criterion $R(-\nabla_F E(\Gamma) + v)$:

$$\hat{v} = \underset{v \perp \nabla_F E(\Gamma)}{\arg\min} \ R(-\nabla_F E(\Gamma) + v) \tag{7}$$

In fact this remark still stands when the choice of the direction of descent is not the gradient itself. If we denote by $u$ the initially proposed deformation field $\frac{d\Gamma}{dt}$, then adding a vector field which is orthogonal to the

gradient $\nabla_F E(\Gamma)$ will not change the amount of decreased energy at this step of the gradient descent (but will change the evolution):

$$\frac{dE(\Gamma)}{dt} = \langle \nabla_F E(\Gamma) | -u + v \rangle_F = \langle \nabla_F E(\Gamma) | -u \rangle_F$$

Note that the notion of being orthogonal to the gradient is independent from the chosen inner product. Indeed, if $F$ and $G$ are two different inner products, $\nabla_F E$ and $\nabla_G E$ the associated gradients, and $\perp_F$ and $\perp_G$ the associated notions of orthogonality, we have:

$$\langle \nabla_F E(\Gamma) | v \rangle_F = \delta E(\Gamma, v) = \langle \nabla_G E(\Gamma) | v \rangle_G$$

so, consequently:

$$\langle \nabla_F E(\Gamma) | v \rangle_F = 0 \iff \langle \nabla_G E(\Gamma) | v \rangle_G = 0$$
$$\nabla_F E(\Gamma) \perp_F v \iff \nabla_G E(\Gamma) \perp_G v \,.$$

# 4. Some Spatially Coherent Minimizing Flows

This theoretical study has brought us the tools we need to better apprehend minimizing flows and build new ones. We now propose some minimizing flows yielding different degrees of spatial coherence. We insist on the fact that this spatial coherence has nothing to do with an eventual regularity term in the energy functional. We do not modify the energy, so the regularity constraint on the contour remains unchanged. We modify the trajectory of the minimizing flow, by favoring spatially coherent motions, but this does not condition the regularity of the final contour.

In the following, we sometimes use differential geometry. We refer the reader to [10] for the basic notions.

## 4.1. Motion decomposition

A simple and useful procedure, to design new inner products yielding spatially coherent flows, is to decompose the deformation space into a sum of several mutually orthogonal linear subspaces, and to apply different penalty factors to the different types of motions. Typically, the subspaces are chosen according to an application-specific hierarchy of the motions. For example, rigid/non-rigid, affine/non-affine, etc.

We suppose that such an orthogonal (with respect to $\langle | \rangle_F$) decomposition of the deformation space $F$ into $N$ closed linear subspaces is available:

$$F = F_1 \perp F_2 \perp \cdots \perp F_N \,.$$

Then a new inner product is derived from $\langle | \rangle_F$ by applying the procedure of Subsection 3.1 with

$$\mathcal{L} = \bigoplus_{i=1}^{N} \lambda_i \operatorname{Id}_{F_i} \,,$$

where $\forall i, \ \lambda_i > 0$. The lower is $\lambda_i$, the shorter is the norm of the velocity fields of subspace $F_i$, and the stronger will be this type of motion in the new gradient flow.

Obviously, $\mathcal{L}$ is symmetric positive definite and invertible. If $\nabla_F E$ exists, so does $\nabla_\mathcal{L} E$ and

$$\nabla_\mathcal{L} E \;=\; \sum_{i=1}^{N} \frac{1}{\lambda_i} \; \Pi_{F_i} \left( \nabla_F E \right) \;, \tag{8}$$

where $\Pi_{F_i}$ denotes the orthogonal projection on the $i^{th}$ subspace $F_i$. Of course, if all $\lambda_i$ are equal to 1, $\nabla_\mathcal{L} E$ coincides with $\nabla_F E$.

We apply this general construction to two useful cases. In the first case, we decompose the velocity field into a translation, an instantaneous rotation, a rescaling motion and a non-rigid residual. In the second case, we isolate the instantaneous affine motion.

In the following, we denote by $\mathbf{G} = \left( \int_\Gamma \mathbf{x} \, d\Gamma(\mathbf{x}) \right) / \int_\Gamma d\Gamma(\mathbf{x})$ the center of mass of the shape.

### 4.1.1 Translation, rotation and scaling

In this paragraph, we focus on the two-dimensional and three-dimensional cases. The expressions below are for the 3D case, but can easily be adapted to 2D.

We denote by $T$, $R$ and $S$ the subspaces of the translations, the instantaneous rotations around the centroid, and the scaling motions centered on the centroid, respectively, defined on the shape $\Gamma$:

$$
\begin{aligned}
T &= \left\{ v : \mathbf{x} \in \Gamma \mapsto \mathbf{t} \mid \mathbf{t} \in \mathbb{R}^3 \right\} \;, \\
R &= \left\{ v : \mathbf{x} \mapsto (\mathbf{x} - \mathbf{G}) \wedge \omega \mid \omega \in \mathbb{R}^3 \right\} \;, \\
S &= \left\{ v : \mathbf{x} \mapsto s(\mathbf{x} - \mathbf{G}) \mid s \in \mathbb{R} \right\} \;.
\end{aligned}
$$

These subspaces are mutually orthogonal for the $L^2$ inner product. Indeed, the $L^2$ product of any two fields of any two different subspaces (among $T$, $R$ and $S$) is zero. For instance, if $v_1 : \mathbf{x} \mapsto \mathbf{t}_1$ is an element of $T$ and $v_2 : \mathbf{x} \mapsto (\mathbf{x} - \mathbf{G}) \wedge \omega_2$ an element of $R$, then:

$$\langle v_1 \,|\, v_2 \rangle_{L^2} \;=\; \frac{1}{|\Gamma|} \int_\Gamma \mathbf{t}_1 \cdot \left( (\mathbf{x} - \mathbf{G}) \wedge \omega_2 \right) d\Gamma(\mathbf{x}) \;=\; \mathbf{t}_1 \cdot \left( \left[ \frac{1}{|\Gamma|} \int_\Gamma \mathbf{x} \, d\Gamma(\mathbf{x}) - \mathbf{G} \right] \wedge \omega_2 \right) \;=\; 0.$$

We suppose that these subspaces are included in the space of admissible deformations $F$, and that the latter is ruled by the $L^2$ inner product. We denote by $N$ the orthogonal complement of these subspaces: $F = T \perp R \perp S \perp N$. The orthogonal projection of a velocity field $u$ on one of these subspaces can be found by minimizing $\|u - v\|_F$ with respect to $v$ in the considered subspace. As an example, we detail the computation of $(\Pi_R u)$.

As for each element $v$ of $R$ there exists an $\omega$ such that $v(\mathbf{x}) = (\mathbf{x} - \mathbf{G}) \wedge \omega$ for all $\mathbf{x}$, we minimize the quantity $\|u - (\cdot - \mathbf{G}) \wedge \omega\|_{L^2}$ with respect to $\omega$.

$$\partial_\omega \left( \int_\Gamma \|u(\mathbf{y}) - (\mathbf{y} - \mathbf{G}) \wedge \omega\|^2 \, d\Gamma(\mathbf{y}) \right) \;=\; \int_\Gamma -\left( u(\mathbf{y}) - (\mathbf{y} - \mathbf{G}) \wedge \omega \right) \wedge (\mathbf{y} - \mathbf{G}) \, d\Gamma(\mathbf{y})$$

$$= -\int_\Gamma u(\mathbf{y}) \wedge (\mathbf{y} - \mathbf{G}) \, d\Gamma(\mathbf{y}) + \left( \int_\Gamma \|\mathbf{y} - \mathbf{G}\|^2 d\Gamma(\mathbf{y}) \right) \omega - \left[ \int_\Gamma (\mathbf{y} - \mathbf{G})(\mathbf{y} - \mathbf{G})^T d\Gamma(\mathbf{y}) \right] \omega$$

As this quantity is zero for the $\omega_u$ which minimizes $\|u - (\cdot - \mathbf{G}) \wedge \omega\|_{L^2}$, we have:

$$\omega_u = \left[ \left( \int_\Gamma \|\mathbf{y} - \mathbf{G}\|^2 d\Gamma(\mathbf{y}) \right) \mathrm{Id} - \int_\Gamma (\mathbf{y} - \mathbf{G})(\mathbf{y} - \mathbf{G})^T d\Gamma(\mathbf{y}) \right]^{-1} \left( \int_\Gamma u(\mathbf{y}) \wedge (\mathbf{y} - \mathbf{G}) \, d\Gamma(\mathbf{y}) \right)$$

To guarantee that the linear application between brackets is invertible, we prove it is a symmetric positive definite matrix $M$. We have indeed for any $\mathbf{x}$:

$$\mathbf{x}^T M \mathbf{x} \;=\; \int_\Gamma \|\mathbf{x}\|^2 \|\mathbf{y} - \mathbf{G}\|^2 - \left(\mathbf{x} \cdot (\mathbf{y} - \mathbf{G})\right)^2 \, d\Gamma(\mathbf{y})$$

As for any $\mathbf{x}$ and $\mathbf{z}$ we have $(\mathbf{x} \cdot \mathbf{z}) \leqslant \|\mathbf{x}\|\|\mathbf{z}\|$, with equality only if the two vectors are collinear, and as $\mathbf{x}$ cannot be collinear with all $\mathbf{y} - \mathbf{G}$ for $\mathbf{y}$ in $\Gamma$, we obtain $\mathbf{x}^T M \mathbf{x} > 0$ for any $\mathbf{x}$, so $M$ is positive definite and consequently invertible.

Note that if we had not taken for $u$ the $L^2$ gradient but the gradient for another inner product $F$, we would have to ensure the subspaces are orthogonal for that inner product $F$, and compute new projections by minimizing $\|u - v\|_F$.

We apply the method we detailed for the subspace $R$ to the other subspaces $T$ and $S$, and obtain:

$$
\begin{aligned}
(\Pi_T\, u)\,(\mathbf{x}) \;&=\; \bar{u} \;:=\; \frac{1}{|\Gamma|} \int_\Gamma u(\mathbf{y})\, d\Gamma(\mathbf{y}), \\
(\Pi_R\, u)\,(\mathbf{x}) \;&=\; (\mathbf{x} - \mathbf{G}) \wedge \omega_u\,, \\
(\Pi_S\, u)\,(\mathbf{x}) \;&=\; \frac{\int_\Gamma u(\mathbf{y}) \cdot (\mathbf{y} - \mathbf{G})\, d\Gamma(\mathbf{y})}{\int_\Gamma \|\mathbf{y} - \mathbf{G}\|^2\, d\Gamma(\mathbf{y})}\, (\mathbf{x} - \mathbf{G})\,, \\
(\Pi_N\, u)\,(\mathbf{x}) \;&=\; u(\mathbf{x}) - (\Pi_T + \Pi_R + \Pi_S)\,(u)(\mathbf{x})\,.
\end{aligned}
$$

In the two-dimensional case, the expressions of the projections are the same, and the expression of $\omega_u$ can be simplified in:

$$\omega_u \;=\; \frac{\int_\Gamma (\mathbf{y} - \mathbf{G}) \wedge u(\mathbf{y})\, d\Gamma(\mathbf{y})}{\int_\Gamma \|\mathbf{y} - \mathbf{G}\|^2\, d\Gamma(\mathbf{y})}\,.$$

The new gradient is deduced from the $L^2$ gradient by equation (5) with

$$\mathcal{L}^{-1} = \mathrm{Id} + \left(\frac{1}{\lambda_T} - 1\right) \Pi_T + \left(\frac{1}{\lambda_R} - 1\right) \Pi_R + \left(\frac{1}{\lambda_S} - 1\right) \Pi_S\,.$$

The weights $\lambda_T$, $\lambda_R$ and $\lambda_S$ are adapted to the user's needs in each particular application. For example:

- Boost rigid+scaling motions: $\lambda_T, \lambda_R, \lambda_S \ll 1$,

- Boost rigid motions: $\lambda_T, \lambda_R \ll 1$, $\lambda_S = 1$,

- Boost translations: $\lambda_T \ll 1$, $\lambda_R = \lambda_S = 1$.

### 4.1.2  Affine motion

We can apply this same idea to the subspace $\mathcal{A}$ of instantaneous affine motions:

$$\mathcal{A} = \left\{ v : \mathbf{x} \in \Gamma \mapsto \mathbf{A}\mathbf{x} + \mathbf{b} \mid \mathbf{A} \in \mathbb{R}^{n \times n},\ \mathbf{b} \in \mathbb{R}^n \right\}\,.$$

The $L^2$ orthogonal projection on this subspace writes:

$$(\Pi_A\, u)\, (\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}\ ,$$

where

$$
\begin{aligned}
\mathbf{A} &= \left[\int_\Gamma u(\mathbf{y})(\mathbf{y}-\mathbf{G})^T\, d\Gamma(\mathbf{y})\right]\left[\int_\Gamma (\mathbf{y}-\mathbf{G})(\mathbf{y}-\mathbf{G})^T\, d\Gamma(\mathbf{y})\right]^{-1}\ ,\\
\mathbf{b} &= \overline{u} - \mathbf{A}\,\mathbf{G}\ .
\end{aligned}
$$

## 4.2. The Sobolev $H^1$ gradient flow

We consider the canonical inner product of the Sobolev space $H^1(\Gamma, \mathbb{R}^n)$ of square integrable velocity fields with square integrable derivatives, defined on the shape $\Gamma$ with values in $\mathbb{R}^n$. For two such fields $u$ and $v$ its expression is:

$$\langle u\, |v\rangle_{H^1} = \frac{1}{|\Gamma|}\int_\Gamma u(\mathbf{x})\cdot v(\mathbf{x})d\Gamma(\mathbf{x})\ +\ \frac{1}{|\Gamma|}\, l^2\int_\Gamma D_\mathbf{x}u(\mathbf{x})\cdot D_\mathbf{x}v(\mathbf{x})d\Gamma(\mathbf{x})\ ,$$

where $D_\mathbf{x}$ denotes the intrinsic derivatives on the contour and $l$ is a characteristic length for the derivation which acts as a weight between the two integrals. The second term of this expression introduces a notion of spatial coherence: not only the length of the velocity field, but also its variations along the contour are penalized. Indeed, $D_\mathbf{x}u(\mathbf{x})$ stands for the matrix of the derivative of the vector field $u$ at the point $\mathbf{x}$ on the manifold $\Gamma$ and consequently expresses how much the field $u$ varies at point $\mathbf{x}$. In the two-dimensional case, $D_\mathbf{x}u(\mathbf{x})$ is simply a vector. In the general case, $D_\mathbf{x}u(\mathbf{x})\cdot D_\mathbf{x}v(\mathbf{x}) = \sum_{i,j}(D_\mathbf{x}u(\mathbf{x}))_{i,j}(D_\mathbf{x}v(\mathbf{x}))_{i,j}$ is the usual inner product between matrices.

By definition of the gradients of $E(\Gamma)$, and then thanks to an integration by parts on the manifold $\Gamma$, we have:

$$
\begin{aligned}
\forall\, v,\quad \langle\nabla_{L^2}E(\Gamma)\,|v\rangle_{L^2} &= \delta E(\Gamma, v) = \langle\nabla_{H^1}E(\Gamma)\,|v\rangle_{H^1}\\
&= \langle\nabla_{H^1}E\,|v\rangle_{L^2} + l^2\,\langle D_\mathbf{x}\nabla_{H^1}E\,|\,D_\mathbf{x}v\,\rangle_{L^2}\\
&= \langle\nabla_{H^1}E - l^2\Delta\nabla_{H^1}E\,\big|\,v\rangle_{L^2}
\end{aligned}
$$

Thus the $H^1$ inner product is related to the $L^2$ inner product as proposed in Subsection 3.1 through the linear operator $\mathcal{L}(u) = u - l^2\Delta u$, where $\Delta$ denotes the intrinsic Laplacian operator on the contour, often called the Laplace-Beltrami operator. As a consequence, the $H^1$ gradient can be obtained from the $L^2$ gradient by solving an intrinsic heat equation with a data attachment term:

$$l^2\,\Delta u = u - \nabla_{L^2}E\ . \tag{9}$$

Interestingly, the solution of equation (9) coincides with

$$\arg\min_u \int_\Gamma \|u(\mathbf{x})\ -\ \nabla_{L^2}E(\Gamma)(\mathbf{x})\|^2\ d\Gamma(\mathbf{x})\ +\ l^2\int_\Gamma \|D_\mathbf{x}u(\mathbf{x})\|^2\ d\Gamma(\mathbf{x}) \tag{10}$$

Intuitively, the $H^1$ gradient is a smoothed version of the $L^2$ gradient and can be obtained by a process similar to the image restoration process on a manifold $\Gamma$, a problem familiar to the image processing community. The factor $l^2$ acts as a parameter balancing the influences of the data term and the regularizing

term. Actually, smoothing a gradient using this particular inner product is a standard "trick", well-known in numerical analysis. As we mentioned previously, this idea has been introduced in computer vision simultaneously by us [8] and by Sundaramoorthi et al. [30]. However, the main point remains that, introducing this smoothing via a modification of the gradient rather than directly from equation (10), warrants that the gradient descent will decrease the energy.

In the two-dimensional case, the shape is a curve which can be parametrized by its arc length $\sigma$, so that any field $u$ defined on $\Gamma$ can be seen as an application from $[0, |\Gamma|]$ to $\mathbb{R}^2$, where $|\Gamma|$ is the length of the curve. The explicit solution of the equation $\Delta u = u - v$ is then known and given by:

$$u(\sigma) \;=\; \frac{1}{2l}\left(e^{\sigma/l}\left(A - \int_0^\sigma e^{-\tau/l}\,v(\tau)\,d\tau\right) \;+\; e^{-\sigma/l}\left(B + \int_0^\sigma e^{\tau/l}\,v(\tau)\,d\tau\right)\right) \tag{11}$$

$$\text{with } A = \frac{e^{|\Gamma|/l}}{e^{|\Gamma|/l} - 1}\;\oint_\Gamma e^{-\tau/l}\,v(\tau)\,d\tau$$

$$\text{and } B = \frac{1}{e^{|\Gamma|/l} - 1}\;\oint_\Gamma e^{\tau/l}\,v(\tau)\,d\tau\,.$$

Of course, the choice of the initial point on $\Gamma$ in order to define its parametrization by the arc length does not interfere with the resulting solution considered as an application from $\Gamma$ into $\mathbb{R}^2$.

In greater dimensions, we can obtain in practice the $H^1$ gradient, solution of equation (9), from an iterative minimization induced by (10). Since the work introduced in [3], implementing a PDE on a surface is affordable in the implicit framework with the level set method [9, 22].

## 4.3. Intrinsic Gaussian smoothing

We apply the procedure of Subsection 3.2 to design a useful minimizing flow: it is a smoothed version of the $L^2$ gradient flow. Hence, to some extent, it resembles the $H^1$ gradient flow of Subsection 4.2. However, here, we apply an *ad hoc* procedure to the $L^2$ gradient without resorting to an inner product.

We define a linear intrinsic smoothing operator which may be seen as the counterpart on the contour of Gaussian smoothing in $\mathbb{R}^{n-1}$, by considering the solution $\tilde{u}$ of the intrinsic heat equation on $\Gamma$ with initial condition $u$:

$$\left\{\begin{array}{l} \tilde{u}(.,0) = u \\ \dfrac{\partial \tilde{u}}{\partial \tau} = \Delta\,\tilde{u} \end{array}\right. , \tag{12}$$

where $\Delta$ denotes the Laplace-Beltrami operator. We then denote by $\mathcal{L}_\tau\,u$ its solution $\tilde{u}(.,\tau)$ at time $\tau \geq 0$.

On the one hand, $\mathcal{L}_\tau$ is symmetric positive. In particular, a flow (6) based on this operator decreases the energy. The larger is $\tau$, the smoother is the flow.

$\mathcal{L}_\tau$ is symmetric:

$$\begin{array}{rclcl} \langle \mathcal{L}_0(u)\,|\,v\rangle_{L^2} & = & \langle u\,|\,\mathcal{L}_0(v)\rangle_{L^2} & = & \langle u\,|\,v\rangle_{L^2}\ , \\[4pt] \dfrac{\partial}{\partial \tau}\langle \mathcal{L}_\tau(u)\,|\,v\rangle_{L^2} & = & \dfrac{\partial}{\partial \tau}\langle u\,|\,\mathcal{L}_\tau(v)\rangle_{L^2} & = & -\langle D_\mathbf{x} u\,|\,D_\mathbf{x} v\rangle_{L^2} \end{array}$$

$\mathcal{L}_\tau$ is positive:

$$\langle \mathcal{L}_\tau(u)\,|\,u\rangle_{L^2} \;=\; \langle \mathcal{L}_{\tau/2}\mathcal{L}_{\tau/2}(u)\,|\,u\rangle_{L^2} = \left\|\mathcal{L}_{\tau/2}(u)\right\|_{L^2} \geq 0$$

But on the other hand, the inversion of $\mathcal{L}_\tau$ for $\tau > 0$ is an ill-posed anti-diffusive process. So a gradient interpretation is not available.

# 5. Numerical Experiments With The New Inner Products

The approach presented in this paper can be applied to virtually any active contour evolution. In this section, we have chosen some particular applications to demonstrate the interest of our contribution.

Moreover, the content of this paper is not specific to a particular implementation of the contour evolution. In our experiments, we have used the level set framework [9, 22, 28, 20, 21], motivated by its numerical stability and its ability to handle topological changes automatically. The implicit framework also offers an elegant formulation of the Laplace-Beltrami operator [2] and of the average of a quantity along the contour [25].

The additional computational cost of our approach depends on the type of minimizing flow we consider. The extra time is barely noticeable for the rigid plus scaling and affine flows of paragraphs 4.1.1 and 4.1.2. The latter only require to compute a handful of integrals on the contour. The smooth minimizing flows of Subsections 4.2 and 4.3 are more demanding. In 2D, the implicit diffusion equations (9) and (12) are equivalent to some convolutions with respect to the curvilinear coordinate on $\Gamma$. In 3D and more, they must be solved with some iterative methods, for each time step.

## 5.1. Shape warping

We illustrate our approach in the problem of shape warping. In this context, the energy functional to be minimized is a measure of dissimilarity between the evolving contour and a target contour. The study of shape metrics is still an active research area [34, 33, 7, 32], and there are many candidates for the dissimilarity measure. In this paper, we use a differentiable approximation of the well-known Hausdorff distance, as proposed in [7], to warp the contours of two different hands.

Figure 1 compares the evolution of the contour when using the $L^2$ gradient descent *(top row)* and a modified gradient descent favoring rigid plus scaling motions *(bottom row)* as in paragraph 4.1.1. Both evolutions achieve a perfect warping. However, despite the similarity of the two input shapes, the $L^2$ gradient flow goes through some states of completely different appearances. The trajectory followed by this flow looks particularly inefficient and unnatural, because the notion of length contained in the $L^2$ inner product is very far from our intuition. In contrast, the behavior of our gradient flow is natural and visually pleasing. Some movies of these evolutions are available in our additional submission data.

In Figure 2, we show a three-dimensional warping example from a teddy bear to Hayao Miyazaki's character Totoro. We use here the $W^{1,2}$-norm of the distance functions as proposed in [7]. Despite an initial rigid registration, the $L^2$ gradient descent is unable to give satisfying results. A modified gradient descent favoring rigid plus scaling motions leads to better results.

This suggests that our approach can infer relevant correspondences between the two contours, as a byproduct of the warping process. This point-to-point matching is obtained by tracking the points along the evolution. It does not make much sense with a $L^2$ gradient flow, because the latter yields a strictly normal velocity field. But when using our approach, the velocity field has a meaningful tangential part. Maintaining point correspondences during the evolution is straightforward in an implementation with meshes. It is also feasible in a level set implementation, with an extension proposed in [26].

## 5.2. Tracking

We now illustrate the better robustness to local minima of spatially coherent minimizing flows with a naive experiment. We insist on the fact that this example is illustrative: we did not look for the method and the energy that gave the best results of tracking for the particular sequence we worked on; we focus more on the improvements brought by our changement of inner product rather than on the results themselves.
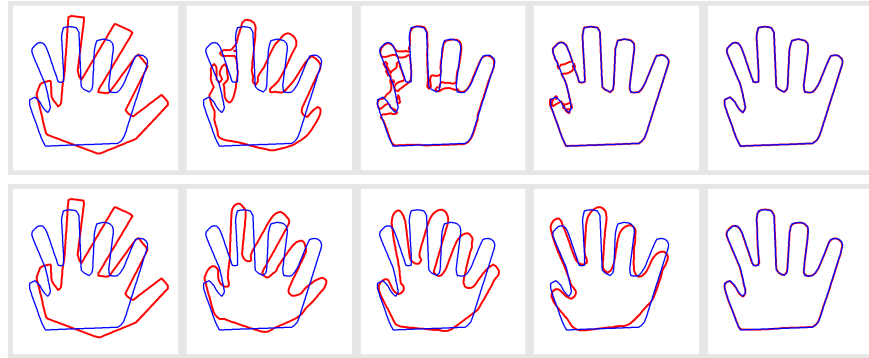
Figure 1: Shape warping with the $L^2$ gradient descent *(top)* and with a modified gradient descent favoring rigid plus scaling motions *(bottom)*: $\lambda_T = \lambda_R = \lambda_S = 0.025$.
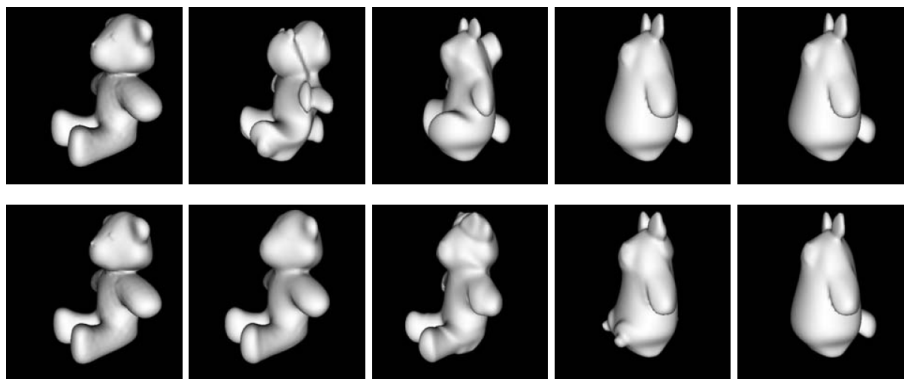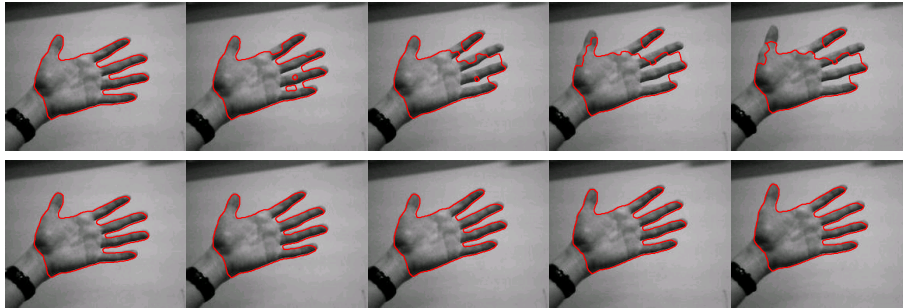


Figure 2: 3D shape warping with the $L^2$ gradient descent *(top)* and with a modified gradient descent favoring rigid plus scaling motions *(bottom)*: $\lambda_T = \lambda_R = \lambda_S = 0.025$.

12

Figure 3: Tracking a hand in a video sequence with the $L^2$ gradient descent *(top)* and with a modified gradient descent favoring affine motions *(bottom)*: $\lambda_A = 0.025$.

We track a moving hand in a monocular video sequence. For each frame, we minimize the contour-based energy of the original geodesic active contours method [6], starting from the result of the segmentation of the previous frame. Note that a region-based approach [24] or a background substraction method would give better results on our particular test sequence.

Figure 3 compares the evolution of the contour when using the $L^2$ gradient descent *(top row)* and a modified gradient descent favoring affine motions *(bottom row)* as in paragraph 4.1.2. Due to large displacements between consecutive frames, the $L^2$ gradient flow fails and the contour finally locks between two fingers, whereas our gradient flow manages to dodge this unwanted low-energy configuration.

## 5.3. Landmarks-guided shape warping

Let us study the case of an energy which does not admit a usual $L^2$ gradient because its formal computation leads to an expression with Dirac peaks in the space of distributions. The problem with such a gradient is that it is not implementable in practice. However, with a suitable choice of another inner product, we naturally obtain a smoothed version of this gradient.

We consider two shapes $\Gamma_1$ and $\Gamma_2$, and we would like to warp $\Gamma_1$ onto $\Gamma_2$. It can be useful to consider a small set of landmarks in order to improve the evolution. Provided by the user (anatomical landmarks), or automatically extracted (geometric landmarks), we assume that we are given $p$ pairs of corresponding points on the initial and on the target shapes, $\{(\mathbf{x}_{1i}, \mathbf{x}_{2i}) \in \Gamma_1 \times \Gamma_2, 1 \leq i \leq p\}$. We would like to use the information given by these correspondences to guide the evolution.

### 5.3.1 Choice of the energy and of the inner product

The usual variational approach consists in minimizing the distance between the evolving shape $\Gamma(t)$ and the target one $\Gamma_2$, with initial condition $\Gamma(0) = \Gamma_1$. This distance $E_d$ could be for example the approximation of the Hausdorff distance presented in [7] or the $W^{1,2}$ norm of the signed distance functions over the embedding space. We would like to add to this distance a landmark term $E_{\mathscr{L}}$; the energy to minimize would be consequently:

$$E(\Gamma, \Gamma_2) = E_d(\Gamma, \Gamma_2) + E_{\mathscr{L}}(\Gamma, \Gamma_2)$$

We follow each landmark $\mathbf{x}_{1i}$ from the initial shape $\Gamma_1$ during the evolution and denote by $\mathbf{x}_i(t)$ its corresponding point on $\Gamma(t)$. We simply choose for the landmark term:

13

$$E_{\mathscr{L}} = \sum_i d(\mathbf{x}_i(t), \mathbf{x}_{2i})^2 \qquad (13)$$

Formally, the energy given by equation (13) yields Dirac peaks in the expression of the gradient of the energy:

$$\forall x \in \Gamma, \ \ \nabla_{L^2} E(\Gamma)(\mathbf{x}) = \nabla_{L^2} E_d(\Gamma)(\mathbf{x}) + \sum_i \delta_{\mathbf{x}_i(t)}(\mathbf{x})(\mathbf{x}_i(t) - \mathbf{x}_{2i}) \qquad (14)$$

where $\delta_{\mathbf{x}}$ denotes the Dirac function centered at point $\mathbf{x}$. This is indeed not a good candidate for a gradient descent.

The trick consists in changing the inner product which appears in the definition of the gradient. We use $H^1(\Gamma, \mathbb{R}^n)$, the Sobolev space of square integrable velocity fields with square integrable derivatives, defined and studied in section 4.2.

Starting from the irregular gradient $\nabla_{L^2} E(\Gamma)$ given by equation (14), we obtain a smooth gradient $\nabla_{H^1} E(\Gamma)$, given by the PDE (9) and mathematically justified by an adapted choice of inner product that guarantees a decrease of the energy. In practice and in detail, when solving (9), we substitute for each Dirac peak in the expression of $\nabla_{L^2} E(\Gamma)$ a Gaussian with a very small standard deviation.

In the two-dimensional case, the equation (11) gives us an explicit expression of the $H^1$ gradient from the $L^2$ one thanks to a convolution. In the three-dimensional case we have to deal with the minimization process proposed in section 4.2.
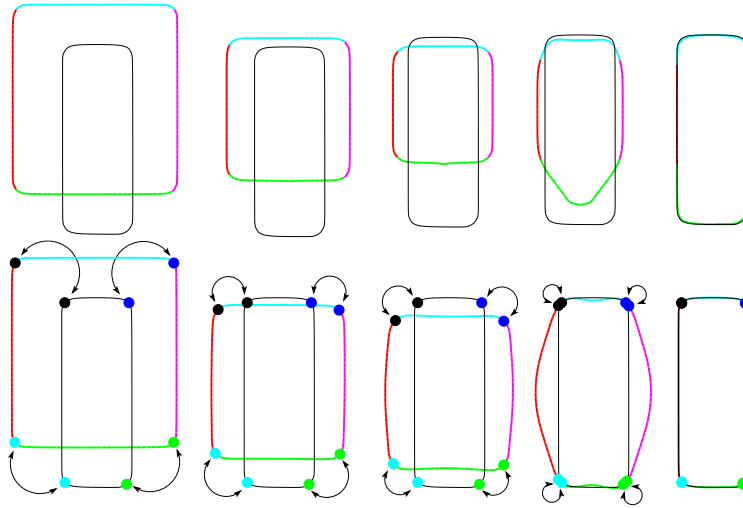
### 5.3.2 Experiments



Figure 4: Warping of a rectangle shape onto another one. Top row: evolution with $E = d_{W^{1,2}}$. Bottom row: evolution with the same energy, augmented with four provided landmarks, marked by color spots. The colors on the evolving curve shows the evolution of different parts of it. See text for comments.
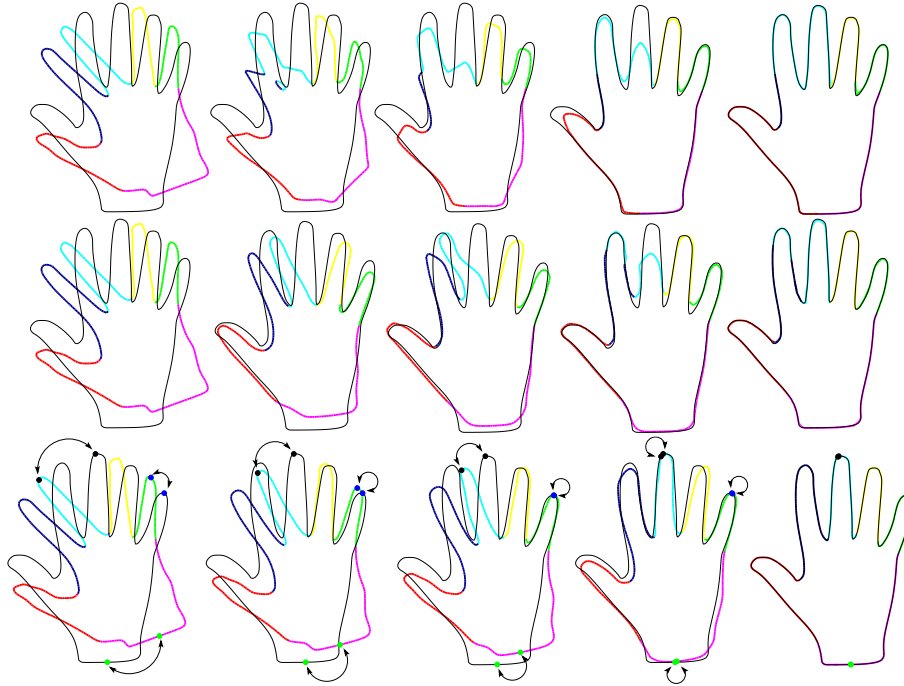
Figure 5: Warping of a hand shape onto another one. Top row: evolution with $E = d_{W^{1,2}}$. Middle row: evolution with the same energy plus spatially coherent flows. Bottom row: evolution with the same energy plus coherent flows plus three provided landmarks. See text for comments.

As a benchmark, we warp some artificial two-dimensional curves with the original energy $E_d = d_{W^{1,2}}$ and test how our landmark-guided force modifies the warping and the final matching. To begin with a simple example, we show in figure 4 the warping of a rectangle onto another one. The different parts of the curves are shown with different colors, so that their respective evolution can be followed. Although the initial warping without any landmark seems natural, it fails discovering the matching between the edges of the rectangles, a matching indeed recovered when providing landmarks. Let us now study the case of some real, complex shapes. Figure 5 shows the warping between two hands. The energy $E = d_{W^{1,2}}$ yields an unnatural warping. Adding spatially coherent flows makes the warping a bit better but still fails in some parts, mainly because the difference between the two shapes can not be summed up to a global motion. With three landmarks only, both a satisfying warping and a good matching are recovered. Figure 6 shows the warping of a teddy bear onto a cartoon character. Without any landmarks, the top row evolution fails matching the ears and arms of the characters. The bottom row shows the evolution with four landmarks. Red spots allow to check a good matching between landmarks.

## 6. A generalized gradient method

In this section, we go further and consider the definition of the gradient of an energy from a new point of view, which leads us to a larger class of minimization algorithms. The thread we follow is the fact that the
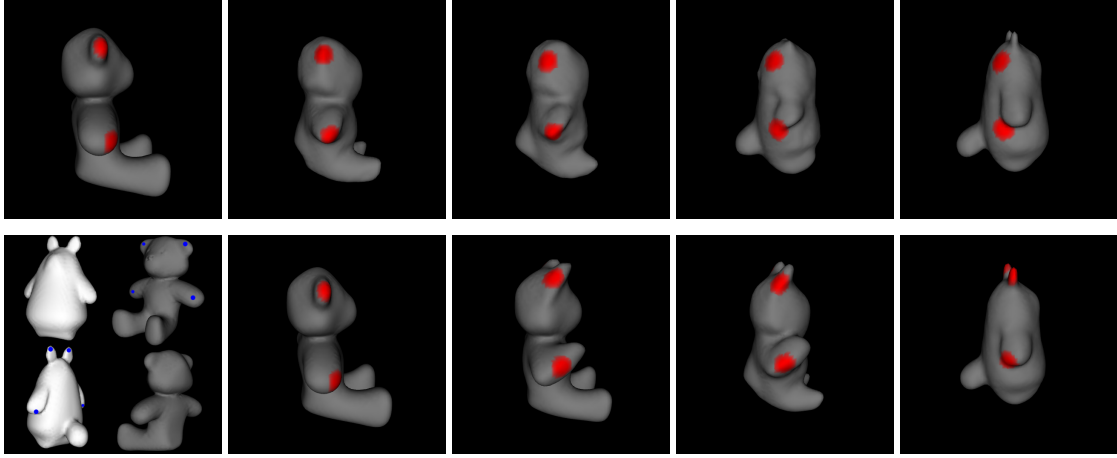
Figure 6: Warping of a teddy bear onto a cartoon character. Top row: evolution with $E = d_{W^{1,2}}$. Bottom row, first image: four landmarks provided on the two shapes, indicated by blue spots. Bottom row, remaining images: evolution with $E = d_{W^{1,2}}$ plus the provided landmarks. In red, some parts of the shapes are tracked. See text for comments.

gradient of the energy can be obtained as the result of another minimization problem.

To help developing the reader's intuition let us recall that the usual gradient descent method can be seen, up to first order, as minimizing $E(\Gamma + u)$ with respect to the deformation field $u$ through the linearization of the energy $E$ in the neighborhood of the shape $\Gamma$:

$$E(\Gamma + u) \simeq E(\Gamma) + \delta E(\Gamma, u)$$

But since $\delta E(\Gamma, u)$ is linear with respect to the deformation field $u$, there is no minimum. This is of course a direct consequence of the first-order approximation. It is therefore more sensible to speak in terms of the direction of the deformation field $u$. The notion of direction implies the choice of a norm: the set of all directions is the set of all fields with norm equal to 1. Once a norm $F$ has been chosen (related to an inner product preferably), a natural solution appears as the direction $u_F$ that minimizes the energy $\delta E(\Gamma, v)$:

$$u_F = \underset{\{v \ \text{s.t.} \ \|v\|_F = 1\}}{\arg\min} [\, \delta E(\Gamma, v) \,] = -\frac{\nabla_F E(\Gamma)}{\|\nabla_F E(\Gamma)\|_F} \tag{15}$$

The main point here is that the opposite of the gradient $-\nabla_F E(\Gamma)$ of the energy $E$ for the inner product related to the norm $F$ is precisely in the direction $u_F$. This gradient has been introduced previously as the deformation field linked to the continuous linear form $\delta E(\Gamma, \cdot)$ for the inner product $F$ thanks to the Riesz theorem. Note that the influence of the inner product $F$ upon the best direction $u_F$ lies in the fact that it changes the shape of the unit sphere (the set of all directions $u$ with unit norm $\|u\|_F = 1$).

It turns out that the gradient itself (not only its direction) can be obtained as the solution of a minimization problem. This also explicits the link between the norm $F$ and the gradient. This is shown in the following

**Theorem 3.** *The gradient $\nabla_F E(\Gamma)$ for the inner product $F$ satisfies:*

$$-\nabla_F E(\Gamma) = \arg\min_v \left[ \delta E(\Gamma, v) + \frac{1}{2}\|v\|_F^2 \right]$$

*Proof.* We have indeed, for any $v$:

$$\begin{aligned} \delta E(\Gamma, v) + \frac{1}{2}\|v\|_F^2 &= \frac{1}{2}\left[\|v\|_F^2 + 2\langle v \,|\, \nabla_F E(\Gamma) \rangle_F \right] \\ &= \frac{1}{2}\left[\|v + \nabla_F E(\Gamma)\|_F^2 - \|\nabla_F E(\Gamma)\|_F^2 \right] \end{aligned}$$

So that:

$$\begin{aligned} \arg\min_v \left[ \delta E(\Gamma, v) + \frac{1}{2}\|v\|_F^2 \right] &= \arg\min_v \left[ \|v + \nabla_F E(\Gamma)\|_F^2 \right] \\ &= -\nabla_F E(\Gamma) \end{aligned}$$

$\square$

The expression between brackets breaks up into two parts: the first one, $\delta E(\Gamma, v)$, comes from the energy $E(\Gamma)$ and stands for the quantity to minimize, whereas the second one, $R(v) = \frac{1}{2}\|v\|_F^2$, is a regularizing term which imposes to the solution to be smooth and small enough in the sense of the norm $F$. Different choices of the smoothing term thanks to different choices of the norm $F$ imply different final gradients $\nabla_F E(\Gamma)$.

For example, the choice of the $H^1$ inner product leads to the regularizing term $R(v) = \frac{1}{2}\|v\|_{L^2}^2 + \frac{1}{2}l^2\|Dv\|_{L^2}^2$ and consequently the gradient $\nabla_{H^1} E(\Gamma)$ is the deformation field which minimizes $\delta E(\Gamma, v) + \frac{1}{2}\|v\|_{L^2}^2 + \frac{1}{2}l^2\|Dv\|_{L^2}^2$. This leads us to an elegant proof of a property of the $H^1$ gradient stated in section 4.2, without considering PDEs:

**Proposition 4.** *The opposite of the $H^1$ gradient is the solution of:*

$$\arg\min_v \left[\|u - v\|_{L^2}^2 + l^2\|Dv\|_{L^2}^2\right]$$

*where $u = -\nabla_{L^2} E(\Gamma)$ is the opposite of the usual gradient.*

*Proof.* Indeed, for any $v$:
$$\|u - v\|_{L^2}^2 = \|u\|_{L^2}^2 - 2\langle u \,|\, v \rangle_{L^2} + \|v\|_{L^2}^2$$

hence
$$\|u - v\|_{L^2}^2 + l^2\|Dv\|_{L^2}^2 = \|u\|_{L^2}^2 + 2\,\delta E(\Gamma, v) + \|v\|_{H^1}^2$$

since by definition of $u$, $\delta E(\Gamma, v) = \langle -u \,|\, v \rangle_{L^2}$; so the $H^1$-gradient can naturally be seen as a smoothed version of the standard gradient $u$, thanks to theorem 3:

$$\arg\min_v \left[\|u - v\|_{L^2}^2 + l^2\|Dv\|_{L^2}^2\right] = \arg\min_v \left[\delta E(\Gamma, v) + \frac{1}{2}\|v\|_{H^1}^2\right] = -\nabla_{H^1} E(\Gamma)$$

$\square$

## 6.1. Generalization of the regularizing term

We have stressed the influence of the choice of an inner product $\langle | \rangle_F$ on the associated gradient:

$$-\nabla_F E(\Gamma) = \arg \min_v [\, \delta E(\Gamma, v) \; + \; R(v) \,]$$

where $R(v) = \frac{1}{2}\|v\|_F^2$, and $\| \cdot \|_F$ is the norm related to the chosen inner product. Since the choice of the inner product is equivalent to the choice of the regularizing term $R(v)$ and acts qualitatively upon the gradient descent paths, we can see $R(v)$ as a prior on the deformation fields.

Let us now generalize our framework and allow $R(v)$ to be (almost) any positive real function, not necessarily related to an inner product, and compute (when it exists) the associated field which we will denote, with a slight abuse of notation, by $-\nabla_R E(\Gamma)$ (note that if $R$ is related as previously to the inner product $F$, then $\nabla_F E = \nabla_R E$):

$$-\nabla_R E(\Gamma) = \arg \min_v [\, \delta E(\Gamma, v) \; + \; R(v) \,] \qquad (16)$$

Under some reasonable assumptions about $R(v)$, the new "gradient" $\nabla_R E(\Gamma)$ exists and has interesting properties.

First, the existence of $\nabla_R E(\Gamma)$, which is the solution of an infimum problem, is guaranteed if $R$ is positive, superlinear and convex. The solution is then not necessarily unique; nevertheless, the set of solutions is convex, reduced to a single point in most cases. However, the question of the existence and unicity of $\nabla_R$ in general is not the main point here, it depends on the particular chosen application $R$. Here, $R$ is supposed to stand for an application approximatively "looking like" the square of a norm; for reasonable choices of $R$ from this point of view, the existence is guaranteed, and so is the uniqueness in most cases.

We now present the fundamental property of the extended gradient $\nabla_R E$ as the

**Theorem 5.** *If $R$ is differentiable and reaches its global minimum at the zero field, then the flow $-\nabla_R E(\Gamma)$, if it exists, decreases the energy $E$.*

*Proof.* We prove that $\delta E(\Gamma, -\nabla_R E(\Gamma)) \leqslant 0$.

We have $-\nabla_R E(\Gamma) = \arg \min_v [\, \delta E(\Gamma, v) \; + \; R(v) \,]$, so, in particular, considering the zero field $v = 0$:

$$\delta E(\Gamma, -\nabla_R E(\Gamma)) \; + \; R(-\nabla_R E(\Gamma)) \; \leqslant \; \delta E(\Gamma, 0) \; + \; R(0)$$
$$\delta E(\Gamma, -\nabla_R E(\Gamma)) \leqslant R(0) - R(-\nabla_R E(\Gamma))$$

As $v = 0$ is the global minimum of $R$, we have $R(-\nabla_R E(\Gamma)) \geqslant R(0)$, so:

$$\delta E(\Gamma, -\nabla_R E(\Gamma)) \leqslant 0$$

Moreover, this last inequality is strict if the usual gradient $\nabla_{L^2} E(\Gamma)$ is not zero. Indeed, in that case, as $R(v)$ reaches its global minimum at the zero field $v = 0$, its derivative is zero for $v = 0$. Consequently, the $L^2$ gradient of $\delta E(\Gamma, v) + R(v)$ with respect to $v$ at the zero field equals $\nabla_{L^2} E(\Gamma)$, which is not zero by hypothesis. Therefore $\inf_v [\, \delta E(\Gamma, v) \; + \; R(v) \,]$ is not reached at $v = 0$ and all inequalities in the proof are strict. $\qquad\square$

Note that the application $R$ is specific to the shape $\Gamma$ (or, more exactly, to its tangent space) and there is no assumption about a "regularity" of the applications $R_\Gamma$ with respect to $\Gamma$. However, as in the previous part where we had to associate to each shape an inner product and naturally chose the same general expression for all of them, we will restrict ourselves to the case where the application $R_\Gamma$ has the same general expression $R(\Gamma)$ for all shapes $\Gamma$ and consequently will commit a slight abuse of notation between $R$ and $R_\Gamma$.

## 6.2. Remarks

### 6.2.1 Addition of an orthogonal term

Note that the method proposed in section 3.3, which consists in adding an orthogonal term to the gradient (see equation (7)), can be seen as a variation on the extended gradient theme, where the search for the infimum has been restricted to the affine hyperplane $\mathcal{H}$ containing the opposite of the gradient $-\nabla_F E$ and orthogonal to it. Indeed:

$$
\begin{aligned}
\arg\min_{v \in \mathcal{H}} \left[ \delta E(\Gamma, v) + R(v) \right] &= \arg\min_{w; w \perp \nabla_F E} \left[ \delta E(\Gamma, -\nabla_F E + w) + R(-\nabla_F E + w) \right] \\
&= \arg\min_{w; w \perp \nabla_F E} R(-\nabla_F E + w)
\end{aligned}
$$

### 6.2.2 Directional formulation

We have seen earlier (equation (15)) that the direction of the gradient could be defined as the field $v$ of the unit sphere $\mathcal{U}_F = \{v \text{ s.t. } \|v\|_F = 1\}$ which most decreases the energy, and that changing the inner product $F$ was precisely acting on the gradient by changing the unit sphere. One way to generalize the notion of gradient could have been to set any hypersurface $\mathcal{S}$ instead of the unit sphere $\mathcal{U}_F$ and to search for the best field $v$ in $\mathcal{S}$. However, this would lead to some difficulties in practice (how to search for a minimum on an hypersurface of infinite dimension, how to represent this hypersurface?). A slightly better way to do this would be to focus on the hypersurfaces of the form $\mathcal{U}_R = \{v \text{ s.t. } R(v) = 1\}$, which is in the spirit of the level-set method. Note that this approach would be very close in practice to the one we described, the main difference being that we would only obtain a direction, without the magnitude.

### 6.2.3 Temporal coherence

The application $R(v)$ does not necessarily only deal with spatial coherence and can also be designed to favor temporally coherent motions. For example, at time step $t$ of an evolution, one could force the new deformation field $u_t$ to resemble the previous one $u_{t-1}$. If we transport $u_{t-1}$ from the previous shape $\Gamma_{t-1}$ to the new one $\Gamma_t$, we obtain a new field noted $T(u_{t-1})$ defined on the same space as $u_t$, and we can compare them, e.g., with $\|T(u_{t-1}) - u_t\|$. We are thus led to define $R(v) = \|T(u_{t-1}) - v\|$. This function however does not satisfy in general the condition $R(0) = 0$ which is necessary in theorem 5. Nevertheless this problem can be solved by defining $R(v)$ to be the norm of the projection of $v$ orthogonally to $T(u_{t-1})$.

## 6.3. Computing the extended gradient

If $R$ is simple enough so that the inverse application of $v \mapsto \nabla_{L^2} R(v)$ is easily computable, then the computation of the extended gradient $\nabla_R E$ is immediate from the knowledge of the usual $L^2$-gradient. Indeed, the application $v \mapsto \delta E(\Gamma, v) + R(v)$ has a local minimum at $v = -\nabla_R E(\Gamma)$, so its derivative with respect to $v$ is zero at this point:

$$
D_v \left( \delta E(\Gamma, v) + R(v) \right)\big|_{v = -\nabla_R E} = 0
$$

$$
\nabla_{L^2} \left( \langle \nabla_{L^2} E(\Gamma) \, | v \rangle_{L^2} + R(v) \right)\big|_{v = -\nabla_R E} = 0
$$

$$
\nabla_{L^2} E(\Gamma) + \nabla_{L^2} R\left( -\nabla_R E(\Gamma) \right) = 0
$$

$$
-\nabla_R E(\Gamma) = (\nabla_{L^2} R)^{-1} \left( -\nabla_{L^2} E(\Gamma) \right)
$$

This formula generalizes the one obtained previously in proposition 2 in section 3 concerning the relation between the gradient for an inner product $P$ and the usual $L^2$ gradient. Now, for the extended gradient, the application $(\nabla_{L^2} R)^{-1}$ which stands in for $L$ in this proposition is not supposed to be linear anymore.

In more general cases, if we cannot compute the application $(\nabla_{L^2} R)^{-1}$, we can still solve the infimum problem with a Partial Differential Equation (PDE) which is equivalent to a ... infinitesimal gradient descent! The definition in equation (16) can be seen indeed as a minimization problem which leads to the evolution:

$$\left\{ \begin{array}{l} v(0) = 0 \\ \dfrac{dv}{dt} = -\nabla_{L^2} E(\Gamma) - \nabla_{L^2} R(v) \end{array} \right. \tag{17}$$

This evolution leads to a local minimum of $\delta E(\Gamma, v) + R(v)$. Even if this local minimum is not the global one (if $R$ has not been well-chosen) or if the evolution is stopped before the convergence, the final flow $v$ computed will strictly decrease the energy $E(\Gamma)$ (same proof as in theorem 5). This point may be important in practice. Note also that there exist many other methods [4] than the usual gradient descent to solve that kind of problem, since the quantity to minimize is a sum of a linear term $\delta E(\Gamma, v)$ and another term $R(v)$ which "looks like" a quadratic term since it is supposed to play a role similar to the square of a norm.

## 6.4. Application: the semi-local rigidification

We now present an example for which the previous framework appears to be useful. We consider an energy $E(\Gamma)$ defined on plane curves. These curves are assumed to lie in an image $\Omega$, in fact a bounded subset of $\mathbb{R}^2$. Instead of following a usual gradient descent in order to minimize $E(\Gamma)$ with respect to $\Gamma$, we would like to favor the deformation fields which preserve the rigidity of the shape as much as possible, or, more exactly, we would like to favor more rigid deformations, so that some kinds of local minima could be avoided. In section 4.1.1 we showed how to change the inner product so as to favor global rigid transformations. In case of articulated movement, this global method may not be sufficient, so we would like to favor fields containing parts close to rigid motions; this leads us to the notion of "semi-local rigidification". We use the expression "semi-local" in order to emphasize the contrast with usual smoothing methods such as Gaussian smoothing or $H^1$ smoothing, which we will qualify of "local".

Let us consider a shape $\Gamma$ and any field $v$ defined on it. We would like to find the parts, if any, of the field $v$ which are well approximated by a translation or a rotation acting on the corresponding parts of $\Gamma$. In order to model this, we associate to each point $\mathbf{x}$ of $\Gamma$ a rigid deformation $w_{\mathbf{x}}$ defined on the whole image $\Omega$. In order to describe $w_{\mathbf{x}}$ we introduce three functions defined on $\Gamma$: a translation $T(\mathbf{x})$, a center of rotation $C(\mathbf{x})$ and the magnitude $A(\mathbf{x})$ of the instantaneous rotation, so that:

$$\forall \mathbf{y} \in \Omega, w_{\mathbf{x}}(\mathbf{y}) = A(\mathbf{x}) \left( \mathbf{y} - C(\mathbf{x}) \right)^{\perp} + T(\mathbf{x})$$

where $a^{\perp}$ stands for the vector $a$ rotated by $+\pi/2$. We suppose that this rigid deformation $w_{\mathbf{x}}$ varies slowly with respect to $\mathbf{x}$, that is to say we suppose the $L^2(\Omega, \mathbb{R}^2)$ norm of its derivative $\|D_{\mathbf{x}} w_{\mathbf{x}}(\cdot)\|_{L^2}$ to be small for each point $\mathbf{x}$ of the curve $\Gamma$. We consider the $L^2(\Gamma, \mathbb{R})$ norm of this application defined on $\Gamma$ and obtain the criterion $\left\| \|D_{\mathbf{x}} w_{\mathbf{x}}(\cdot)\|_{L^2(\Omega, \mathbb{R}^2)} \right\|_{L^2(\Gamma, \mathbb{R})}$ to quantify the smoothness of the field $w_{\mathbf{x}}$ of rigid deformations on $\Gamma$.

It is always possible to express any field $v$ on $\Gamma$ as a member of the class rigid motions:

$$\forall \mathbf{x} \in \Gamma, v(\mathbf{x}) = w_{\mathbf{x}}(\mathbf{x}) = A(\mathbf{x})(\mathbf{x} - C(\mathbf{x}))^{\perp} + T(\mathbf{x}) \tag{18}$$

The field $v$ is then completely defined by the knowledge of $T$, $A$ and $C$. For a given field $v$, there exist of course many triplets $(T, A, C)$ satisfying (18), the simplest one being $(v, 0, G)$, where $G_\Omega$ is the center of mass of the image $\Omega$. In order to lift this ambiguity we define a deformation prior $R$ which depends on $T$, $A$ and $C$ that should be seen as parameters of $v$:

$$R(T, A, C) = \|v\|_{L^2}^2 + \left\| \|D_\mathbf{x} w_\mathbf{x}(\cdot)\|_{L^2(\Omega, \mathbb{R}^2)} \right\|_{L^2}^2$$

which in fact can also be written simpler (by expanding and integrating the expression $\|D_\mathbf{x} w_\mathbf{x}(\mathbf{y})\|^2$) as:

$$R(T, A, C) = \|v\|_{L^2}^2 + \left\| DT + DA\,(G_\Omega - C)^\perp - A\,DC^\perp \right\|_{L^2}^2 + \sigma_\Omega^2 \|DA\|_{L^2}$$

where $\sigma_\Omega^2 = \int_\Omega (\mathbf{y} - G_\Omega)^2 d\mathbf{y}$ is a characteristic squared "length" of the image. The middle term represents the interaction between $T$, $A$ and $C$; for example, changing the center of rotation $DC(\mathbf{x})$ has no consequence on the rigid motion $w_\mathbf{x}$ if it is compensated by the adequate added translation $DT = A\,DC^\perp$. Note that the quantities $G_\Omega$ and $\sigma_\Omega$ are the only ones where the influence of the image $\Omega$ appears.

In order to compute the generalized gradient $\nabla_R$ of an energy $E(\Gamma)$, we first compute the usual $L^2$ gradient $\nabla_{L^2} E$, initialize $(T, A, C) = (0, 0, G_\Omega)$ so that the corresponding field $v(T, A, C)$ is zero, as required in equation (17), and let $(T, A, C)$ evolve to minimize $\delta E(\Gamma, v) + R(T, A, C)$. The corresponding PDEs are

$$\begin{cases} \partial_t T(\mathbf{x}) &=& -\big(\nabla_{L^2} E(\Gamma)(\mathbf{x}) + v(\mathbf{x})\big) \; + \; \Delta \overline{w_\mathbf{x}} \\ \partial_t A(\mathbf{x}) &=& -\big(\nabla_{L^2} E(\Gamma)(\mathbf{x}) + v(\mathbf{x})\big) \cdot \big(\mathbf{x} - C(\mathbf{x})\big)^\perp \\ && + \; \big(G_\Omega - C(\mathbf{x})\big)^\perp \cdot \Delta \overline{w_\mathbf{x}} \; + \; \sigma_\Omega^2 \Delta A(\mathbf{x}) \\ \partial_t C(\mathbf{x}) &=& -A(\mathbf{x})\big(\nabla_{L^2} E(\Gamma)(\mathbf{x}) + v(\mathbf{x})\big)^\perp \; + \; A(\mathbf{x}) \Delta \overline{w_\mathbf{x}}^\perp \end{cases}$$

where $\overline{w_\mathbf{x}} = w_\mathbf{x}(G_\Omega)$ is the mean of the linear application $\mathbf{y} \mapsto w_\mathbf{x}(\mathbf{y})$ on $\Omega$. Note that if we had considered only translations $T$ (and not rotations), we would have $\overline{w_\mathbf{x}} = T(\mathbf{x}) = v(\mathbf{x})$ and the algorithms would act as an $H^1$ smoothing.

## 6.5. Numerical Example

We now apply this method to a specific choice of the energy $E$ to minimize. We would like to warp a given initial shape $\Gamma_1$ onto a given target shape $\Gamma_2$, that is to say, we would like to minimize the shape distance between $\Gamma_1$ and $\Gamma_2$ with respect to $\Gamma_1$. We choose for $E$ a smooth approximation of the Hausdorff distance described in [7], which we will denote here by $d_H(\Gamma_1, \Gamma_2)$.

This energy $E$ achieves generally good warping between any two shapes which are relatively close one to the other, but, in case of large deformations, it can suffer from an important lack of spatial coherence if a part $A$ of the moving shape $\Gamma_1$ has to cross a part $B$ of the target one on its way to another parallel part $C$ of the target shape (see figure 7 for an example), because the part $A$ tries to minimize its distance to both parts $B$ and $C$ at the same time.

A global coherence can nonetheless be recovered by an adequate change of inner product which favors rigid transformations, as presented before. However, this is not sufficient for dealing with local deformations. The methods of Gaussian or $H^1$ smoothing studied in sections 4.3 and 4.2 could be helpful, since their action is local. But even if their influence is appreciable, these smoothing techniques do not favor semi-locally rigid deformations like the movements of an articulated object. We have noticed that, in practice, the quality of the matching between two shapes $\Gamma_1$ and $\Gamma_2$ generally depends on the quality of the path that has been followed during the evolution from one shape to the other, or, more precisely, on how natural a human observer would
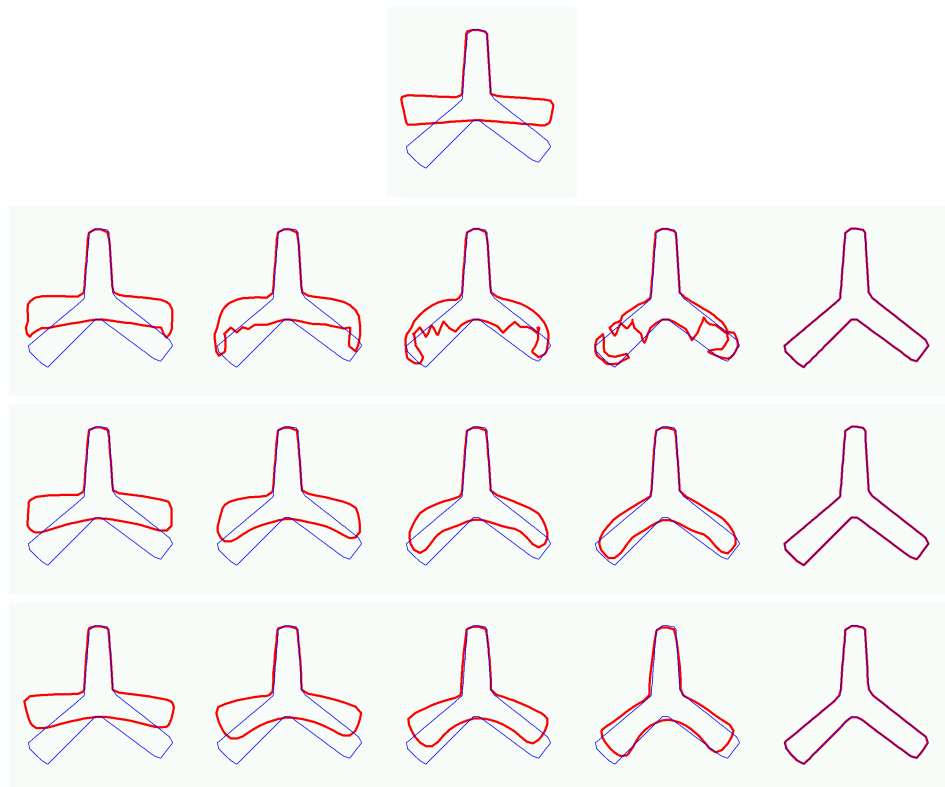
Figure 7: Warping the red shape onto the blue one *(top)* with the $L^2$ gradient descent *(first row)*, with a $H^1$ gradient descent *(second row)* and with a modified gradient descent favoring semi-local rigid motion *(third row)* for the same energy (Hausdorff distance). All evolutions converge to the same shape, but with different paths.
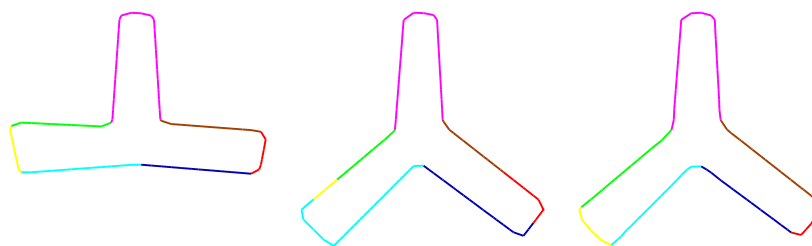


Figure 8: Comparison of the correspondences between the initial curve *(left)* and two final curves resulting from the $H^1$ gradient evolution *(middle)* and from the semi-local rigidification *(right)*. The different parts of the curves are shown with different colors, so that their respective evolutions can be followed. The correspondences for the semi-local rigification case are more geometrically meaningful.
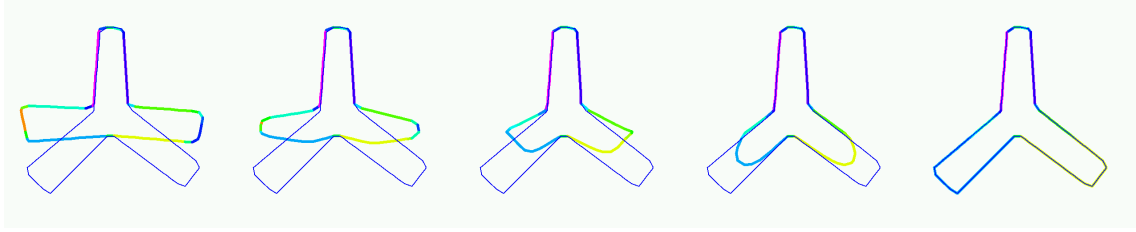
Figure 9: Comparison with a $L^2$ gradient descent on the $L^2$ norm of the signed distance functions associated to the curves. The gradient is naturally smooth but lacks geometric information.

judge this path. This statement is very intuitive and qualitative but we believe that this quality precisely relies on notions such as articulated motion. There is clearly room here for further work. In any case this is the reason why we think that methods like the ones proposed in this paper, which allow to set priors on the deformation fields can have interesting practical applications.

We use the framework presented above and compare the evolutions resulting from three different approaches on a difficult example of shape warping in the case of the Hausdorff distance: the usual $L^2$ gradient method, the $H^1$ smoothing method of section 4.2 (for the best value of the smoothness parameter $l$ in equation (10)) and the semi-local rigidification method (Figure 7). The last one achieves the best path and the best correspondences (Figure 8). Some movies of these evolutions are available in our additional submission data.

The gradient descent framework in the case of an extended gradient $\nabla_R$ could have needed some important additional time if we had to wait until the evolution of $\nabla_R$ converges at each time step of the global evolution of $\Gamma_1$. Fortunately, when necessary, thanks to the remark in section 6.3, we can choose to stop the evolution of $\nabla_R$ before convergence in order to keep the additional cost into reasonable limits. The result presented here was computed so that the total evolution time was multiplied by two, but the effect of the semi-local rigidification is already noticeable for an added cost time of 10%.

For the particular example presented in figure 7, one could object we should have considered other distances, such as the $L^2$ norm between the signed distance functions of the shapes, which always leads to very smooth evolutions. However, those smooth evolutions are not very sensible, in that they often miss entirely the similarity between the two shapes to match (see figure 9). As a consequence their gradient does not contain a lot of geometric information and cannot be very much improved by changes of inner products. This is why, despite the sometimes irregular behavior of the gradient of the Hausdorff distance, we prefer to use it in combination with new inner products, because this has both advantages of guaranteeing smoothness and making geometric sense.

In figure 10 we show an example with real contours from hand segmentation of pictures. As in the previous example, we show the evolution path obtained by minimization of the approximation of the Hausdorff distance between the two curves, with the semi-local rigidification approach. The evolution mainly consists in four local rotations (arms and legs), which fits well our intuition. We have colored, as previously, different parts of the initial curve in order to follow them through the evolution and notice how relevant the correspondences are. A usual gradient descent for this energy would have faced the same irregularity problems as in the first evolution presented in figure 7, and the choice of other usual energies, like the $L^2$ norm between the signed distance functions to the curves, would lack geometric sense, as in figure 9.
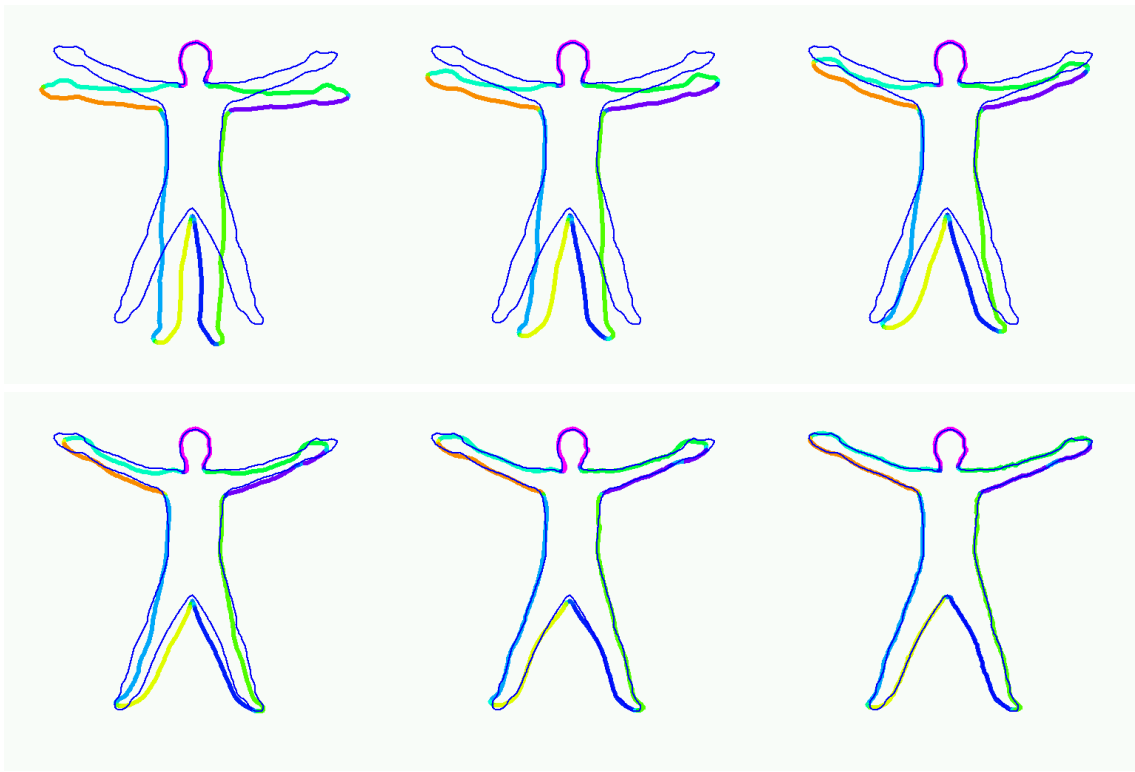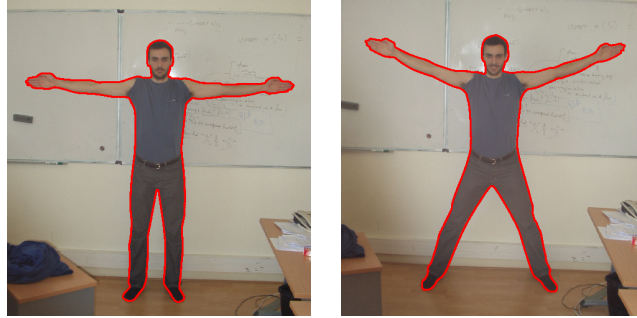
Figure 10: Warping real contours by minimization of the approximation of the Hausdorff distance with the semi-local rigidification approach. The colors show the correspondences between the moving curve and the initial one.

# 7. Conclusion

The impact of the inner product structure of the deformation space on the behavior of the active contours method had been overlooked so far in the computer vision community. We have explored several families of inner products, as well as some minimizing flows not deriving from any inner product by extending the notion of gradient. Given an energy, we now have several ways to minimize it, each of the proposed flows being a minimizing flow but leading to different kinds of evolutions. The inner products and the extended gradients should consequently be seen as priors on the deformation fields, that is to say priors on the evolution paths. They can be used to introduce different degrees of spatial coherence (local, semi-local or global) in the evolution of the contour.

We have shown, with some numerical experiments, that these evolutions better fit our intuitive notion of deformation cost and that they can mimic the behavior of the objects of interest. As a result, they are at the same time more meaningful and more robust to local minima attraction.

# A. Brief Overview of Shape Metrics

For more details on this part, we refer to [7].

## A.1. The Hausdorff distance

Let $\Gamma_1$ and $\Gamma_2$ be two shapes, i.e. manifolds embedded in $\mathbb{R}^n$. The Hausdorff distance between $\Gamma_1$ and $\Gamma_2$ is:

$$
\begin{aligned}
d_H(\Gamma_1, \Gamma_2) &= \sup\left(\sup_{\mathbf{x}\in\Gamma_1}\inf_{\mathbf{y}\in\Gamma_2} d(\mathbf{x},\mathbf{y}), \ \sup_{\mathbf{y}\in\Gamma_2}\inf_{\mathbf{x}\in\Gamma_1} d(\mathbf{x},\mathbf{y})\right) \\
&= \sup\left(\sup_{\mathbf{x}\in\Gamma_1} d(\mathbf{x},\Gamma_2), \ \sup_{\mathbf{y}\in\Gamma_2} d(\mathbf{y},\Gamma_1)\right)
\end{aligned}
$$

The Hausdorff distance is a natural distance on the set of shapes and fits well enough our intuition of distances between shapes, in that, if you consider a (shortest) path $\Gamma(t)$ in the set of shapes between $\Gamma(0) = \Gamma_1$ and $\Gamma(1) = \Gamma_2$ and follow the points of the initial shape during the evolution, then $d_H(\Gamma_1, \Gamma_2)$ is precisely the maximum distance that has to be covered by one of the points to reach its final position on the curve $\Gamma_2$.

However (and consequently), the Hausdorff distance does not handle with considerations such as weak costs for rigid motion; hence our approach consististing in changing the inner product so as to enhance rigid motion.

The Hausdorff distance being not differentiable, we have to consider a smooth approximation of it. We first introduce the following notation:

$$
\langle f \rangle_{\Gamma_1}^{\varphi} = \varphi^{-1}\left(\frac{1}{\Gamma_1}\int_{\mathbf{x}\in\Gamma_1}\varphi(f(\mathbf{x}))\, d\Gamma_1(\mathbf{x})\right)
$$

where $\varphi$ is an application from $\mathbb{R}^+$ to $\mathbb{R}^+$. If $\varphi$ is smooth and increases quickly, then $\langle f \rangle_{\Gamma_1}^{\varphi}$ is a smooth approximation of $\sup_{\Gamma_1} f$. On the other side, if $\varphi$ is smooth and decreases quickly, it is a smooth approximation of the infimum. For well-chosen applications $\varphi$, $\psi$ and $\Psi$, we have found our smooth approximation of the Hausdorff distance:

$$
\rho_H(\Gamma_1, \Gamma_2) = \left\langle \left\langle \langle d(\cdot,\cdot)\rangle_{\Gamma_2}^{\varphi}\right\rangle_{\Gamma_1}^{\psi}, \ \left\langle \langle d(\cdot,\cdot)\rangle_{\Gamma_1}^{\varphi}\right\rangle_{\Gamma_2}^{\psi}\right\rangle^{\Psi}
$$

where $\langle a,\ b\rangle^{\Psi} \ = \ \Psi^{-1}\left(\dfrac{1}{2}(\Psi(a) + \Psi(b))\right)$ is the discrete case of the approximation.

## A.2. Metrics based on distance functions

We associate to any closed shape $\Gamma_1$ its signed distance function $d_{\Gamma_1}$ defined on the whole embedding space $\mathbb{R}^n$, with negative values in the bounded part of $\mathbb{R}^n$ of boundary $\Gamma_1$:

$$d_{\Gamma_1}(\mathbf{x}) \ = \ \begin{cases} \ -d(\mathbf{x}, \Gamma_1) & \text{if } \mathbf{x} \text{ is inside } \Gamma_1, \\ \ +d(\mathbf{x}, \Gamma_1) & \text{else.} \end{cases}$$

We then consider two natural distances on the set of shapes based on their representation by signed distance functions.

$$d_{L^2}(\Gamma_1, \Gamma_2) \ = \ \|d_{\Gamma_1} - d_{\Gamma_2}\|_{L^2(\mathbb{R}^n \mapsto \mathbb{R})} \ = \ \left(\int_{\mathbf{x} \in \mathbb{R}^n} |d_{\Gamma_1}(\mathbf{x}) - d_{\Gamma_2}(\mathbf{x})|^2 \, d\mathbf{x}\right)^{1/2}$$

$$\begin{aligned} d_{W^{1,2}}(\Gamma_1, \Gamma_2) \ &= \ \|d_{\Gamma_1} - d_{\Gamma_2}\|_{W^{1,2}(\mathbb{R}^n \mapsto \mathbb{R})} \\ &= \ \left(\int_{\mathbf{x} \in \mathbb{R}^n} |d_{\Gamma_1}(\mathbf{x}) - d_{\Gamma_2}(\mathbf{x})|^2 \ + \ \|\nabla d_{\Gamma_1}(\mathbf{x}) - \nabla d_{\Gamma_2}(\mathbf{x})\|^2 \ d\mathbf{x}\right)^{1/2} \end{aligned}$$

The minimization of these distances with respect to the shape $\Gamma_1$ leads to naturally smooth evolutions. However, these evolutions may lack geometrical sense in some cases.

## References

[1] N.I. Akhiezer and I.M. Glazman. *Theory of Linear Operators in Hilbert Space*. Pitman, 1981.

[2] M. Bertalmío, L.T. Cheng, S. Osher, and G. Sapiro. Variational problems and partial differential equations on implicit surfaces. *Journal of Computational Physics*, 174(2):759–780, 2001.

[3] M. Bertalmio, G. Sapiro, L.T Cheng, and S. Osher. Variational problems and PDE's on implicit surfaces. In IEEE, editor, *IEEE Workshop on Variational and Level Set Methods*, pages 186–193, Vancouver, Canada, July 2001.

[4] J.F. Bonnans, J.C. Gilbert, C. Lemarechal, and C.A. Sagastizabal. *Numerical Optimization: Theoretical and Practical Aspects*. Springer-Verlag, 2002.

[5] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *International Conference on Computer Vision*, volume 1, pages 26–33, 2003.

[6] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *The International Journal of Computer Vision*, 22(1):61–79, 1997.

[7] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations of Computational Mathematics*, 5(1):1–58, February 2005.

[8] G. Charpiat, R. Keriven, J.P. Pons, and O. Faugeras. Designing spatially coherent minimizing flows for variational problems based on active contours. In *10th International Conference on Computer Vision*, Beijing, China, 2005.

[9] A. Dervieux and F. Thomasset. A finite element method for the simulation of Rayleigh-Taylor instability. *Lecture Notes in Mathematics*, 771:145–159, 1979.

[10] M. P. DoCarmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976.

[11] Y. Duan, L. Yang, H. Qin, and D. Samaras. Shape reconstruction from 3D and 2D data using PDE-based deformable surfaces. In *European Conference on Computer Vision*, volume 3, pages 238–251, 2004.

[12] O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, 1998.

[13] B. Goldlücke and M. Magnor. Space-time isosurface evolution for temporally coherent 3D reconstruction. In *International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 350–355, 2004.

[14] B. Goldlücke and M. Magnor. Weighted minimal hypersurfaces and their applications in computer vision. In *European Conference on Computer Vision*, volume 2, pages 366–378, 2004.

[15] H. Jin, S. Soatto, and A.J. Yezzi. Multi-view stereo beyond Lambert. In *International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 171–178, 2003.

[16] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *The International Journal of Computer Vision*, 1(4):321–331, 1987.

[17] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *European Conference on Computer Vision*, volume 3, pages 82–96, 2002.

[18] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

[19] P.W. Michor and D. Mumford. Riemannian geometries of space of plane curves. Preprint, 2005.

[20] S. Osher and R. Fedkiw. *The Level Set Method and Dynamic Implicit Surfaces*. Springer-Verlag, 2002.

[21] S. Osher and N. Paragios, editors. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer Verlag, 2003.

[22] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[23] N.C. Overgaard and J.E. Solem. An analysis of variational alignment of curves in images. In *International Conference on Scale Space and PDE Methods in Computer Vision*, pages 480–491, 2005.

[24] N. Paragios and R. Deriche. Geodesic active regions and level set methods for motion estimation and tracking. *Computer Vision and Image Understanding*, 97(3):259–282, 2005.

[25] D. Peng, B. Merriman, S. Osher, H.-K. Zhao, and M. Kang. A PDE-based fast local level set method. *Journal of Computational Physics*, 155(2):410–438, 1999.

[26] J.-P. Pons, G. Hermosillo, R. Keriven, and O. Faugeras. How to deal with point correspondences and tangential velocities in the level set framework. In *International Conference on Computer Vision*, volume 2, pages 894–899, 2003.

[27] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1966.

[28] J.A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Sciences*. Cambridge Monograph on Applied and Computational Mathematics. Cambridge University Press, 1999.

[29] J.E. Solem and N.C. Overgaard. A geometric formulation of gradient descent for variational problems with moving surfaces. In *International Conference on Scale Space and PDE Methods in Computer Vision*, pages 419–430, 2005.

[30] G. Sundaramoorthi, A.J. Yezzi, and A. Mennucci. Sobolev active contours. In *IEEE Workshop on Variational and Level Set Methods*, pages 109–120, Beijing, China, 2005.

[31] A. Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *The International Journal of Computer Vision*, 28(3):213–21, 1998.

[32] A.J. Yezzi and A.C.G. Mennucci. Metrics in the space of curves. Preprint, 2005.

[33] A.J. Yezzi and S. Soatto. Deformotion: Deforming motion, shape average and the joint registration and approximation of structures in images. *The International Journal of Computer Vision*, 53(2):153–167, 2003.

[34] L. Younes. Computable elastic distances between shapes. *SIAM Journal of Applied Mathematics*, 58(2):565–586, 1998.

[35] H.-K. Zhao, S. Osher, B. Merriman, and M. Kang. Implicit and non-parametric shape reconstruction from unorganized points using a variational level set method. *Computer Vision and Image Understanding*, 80(3):295–314, 2000.