

# Deformations, Warping and Object Comparison

## A tutorial

Laurent Younes  
CMLA, ENS de Cachan  
61, avenue du président Wilson 94 235 Cachan CEDEX  
e-mail: younes@cmla.ens-cachan.fr  
homepage: www.cmla.ens-cachan.fr/~younes

July 7, 2000

### Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The size of diffeomorphisms</b>	<b>4</b>
2.1	Definitions . . . . .	4
2.2	Standard functional norms . . . . .	6
2.3	Hyperelastic models . . . . .	8
2.4	Iterating small deformations: geodesic energies . . . . .	9
2.5	Learning energies from data-sets . . . . .	10
2.5.1	Generalities . . . . .	10
2.5.2	Implementation . . . . .	12
2.5.3	Learning deformation energies . . . . .	13
<b>3</b>	<b>Landmark-based matching</b>	<b>13</b>
3.1	Splines . . . . .	14
3.2	Bookstein's splines . . . . .	16
3.3	Joshi's interpolation . . . . .	18
<b>4</b>	<b>Dense matching</b>	<b>20</b>
4.1	Variational approaches . . . . .	22
4.1.1	1D Matching . . . . .	22
4.1.2	Image and volume matching . . . . .	25
4.1.3	Numerical implementation . . . . .	29

<b>5</b>	<b>Defining distances</b>	<b>35</b>
5.1	General facts . . . . .	35
5.2	Infinitesimal approach . . . . .	36
5.3	Invariant distances between landmarks . . . . .	37
5.4	Invariant distances on groups of diffeomorphisms . . . . .	40
5.5	Mixing deformations and object variations: landmark matching	41
5.6	Mixing deformations and object variations: image matching .	42

## 1 Introduction

When trying to compare objects which have different natures an essential step consists in finding a common reference scale or common coordinates to quantify them consistently. Most of the methods developed in this presentation are devoted to this purpose: compare visual objects after having found a good common reference system to represent them; then, propose some evaluation framework to quantify the discrepancies.

Probably, the simplest visual objects are families of labeled points, or *landmarks*. They are ordered series of locations (in  $\mathbb{R}^2$  or  $\mathbb{R}^3$ ), each of them being associated to a specific, identifiable, feature of a surface or a volume. Typical examples are face images (cf figure 1), in which landmarks may represent the corner of the eyes, or of the mouth, the eyebrows, etc... It is a rapidly obvious fact in experiments that the family of landmarks acquired from different images or volumes are measured with different coordinate systems, and can only be compared after being placed on a common reference frame. In most applications, two sets of landmarks which can be deduced one from another by a combination of a translation, rotation or scaling (affine similitudes) have to be considered as equal. Comparison can only be made after correctly registering them on a common grid, which must be correctly selected.

For this purpose, Procrustean analysis starts from a quite natural suggestion: select the most favorable grid, the one in which the landmarks are closest. A rigorous formalization of this approach led to the beautiful theory of shape initiated by D. Kendall and his collaborators, in which the manifolds on which sets of landmarks of given cardinality lie are rigorously studied, together with statistical tools designed for their analyses. A short incursion into this theory will be the object of section 5.3, during our discussion on the ways to define distances between objects like shapes.

However, our main interest concerns deformations more general than affine similitudes, which can also be accounted for while comparing sets



Figure 1: Landmarks on a face (image taken from the Olivetti face database)

of landmarks. It is indeed quite common that objects have to be considered as close although they cannot be aligned by any affine transformation. However, general deformations have under mild conditions the ability to transform any set of landmarks into any other one, which implies in particular that the Procrustean approach can no longer be applied. Introducing non-rigid variations requires the ability to quantify the size of the deformation: once done, one would say that two sets of landmarks are close if there exists a small deformation which places them in similar positions.

We shall explore various approaches to quantify deformations. From a general point of view, one already can exhibit two classes: one which is based on prior modeling of the image, and another which is based on a statistical learning of the variations in shape. The first class is loosely related to elasticity theory (or sometimes fluid mechanics), and uses universal, context independent, definitions of when deformations have to be considered as large. The second point of view quantifies large deformations after performing statistical analyses of observed variations within sets of landmarks which should be considered as similar; deformations which might have been considered as large from the first point of view may therefore be neglected from the second one, because they are frequent in the considered class of landmarks, and thus considered as non-significant. The latter approach, initiated by Cootes et al. ([20]), is clearly more appealing, but requires the observation of a large enough number of data to be able to make reliable

statistical conclusions. This will be addressed in section 2.5.

Sets of landmarks are indeed not the only types of objects which can be compared in computer vision. Dealing with 1D objects, like curves, (outlines of shapes), or 2D, like images of surfaces, or even 3D objects is routine. Beside the non-negligible (and often dramatical) increasing of data size, the complication induced by considering “continuous” objects is that, in most cases, one cannot assume anymore that a labeling information is available, in the sense that one does not explicitly know which part of an object should be compared to a given part of the other. In fact, a matching problem appears as an unavoidable part of comparison, leading to the generic class of “elastic matching methods”, on which we essentially focus.

The rest of these notes is organized as follows. We first discuss how deformations can be defined between objects and how their sizes can be evaluated. We then discuss comparison methods, based on variational formulations, placing them under the category of “elastic matching”, together with some algorithms designed for this purpose. We then focus on the problem of finding true distances between objects, show how this can be plugged into Kendall’s theory of shape, in the case of rigid deformations, and how this leads to formulations reminiscent of fluid mechanics in a high dimensional context.

## 2 The size of diffeomorphisms

### 2.1 Definitions

In this section, we provide various ways to quantify variations due to an alteration of an object caused by the action of a diffeomorphism.

We shall mainly consider objects of two kinds:

- Landmarks. These are ordered collections of points in Euclidean space. More precisely, let  $\Omega \subset \mathbb{R}^k$ : the set  $\mathcal{L}^N(\Omega)$  is by definition the set  $\Omega^N$ , composed with all  $N$ -uples  $(x_1, \dots, x_N)$ , with  $x_i \in \Omega$ . Thus  $\mathcal{L}^N(\Omega)$  is the set of all collections of  $N$  landmarks in  $\Omega$ .
- Functions. Formally they consist in mappings  $I : \Omega \rightarrow \mathbb{R}^d$  from the open set  $\Omega \subset \mathbb{R}^k$  into  $\mathbb{R}^d$ , possibly subject to boundary or smoothness conditions. We denote  $\mathcal{F}_d(\Omega)$  the set of such functions. We shall also speak of *images*, by analogy with the 2D case, even if  $k \neq 2$

Our purpose is essentially to compare elements of  $\mathcal{L}^N(\Omega)$ , or  $\mathcal{F}_d(\Omega)$  for a fixed  $\Omega$ . For computer vision applications, elements of  $\mathcal{F}_d(\Omega)$  should be

considered as *parametrized visual objects*. In this context, we essentially consider methods which permit to compare objects which are parametrized by the same set  $\Omega$ . Very often, the choice of  $\Omega$  is quite natural: images can be represented as mappings from a rectangle in  $\mathbb{R}^k$  to  $\mathbb{R}$  (grey-levels) or  $\mathbb{R}^3$  (colors). Curves can be parametrized by arc-length, and represented by their curvature functions, or their tangents. Sometimes, this choice is less clear: plane representations of faces have varying boundaries, even for a single individual, so that the interior grey-level variations are not naturally defined on a unique set  $\Omega$ , independent of the face. However, it is possible in this case to choose a rough, invariant  $\Omega$  (some kind of ellipse) which will provide a reasonable approximation of the face contours.

With these definitions, we refer to  $\Omega$  as the *background space*. For functions,  $\mathbb{R}^d$  is the *value space*, and for landmarks, the value space is the set of labels, namely  $\{1, \dots, N\}$  in our formulation. We also define  $\mathcal{H}(\Omega)$ , the set of homeomorphisms on  $\Omega$ , that is the set of mappings  $\Omega \rightarrow \Omega$  which are continuous and have a continuous inverse. This is a group, in the sense that, if  $g, h \in \mathcal{H}(\Omega)$ , then  $g \circ h^{-1} \in \mathcal{H}(\Omega)$ . A diffeomorphism is a differentiable homeomorphism with differentiable inverse. In the following, we will be considering subgroups  $G$  of  $\mathcal{H}(\Omega)$  subject to various constraints of smoothness and boundary conditions. The definition of such groups is by no means an obvious matter, but we shall not enter into these details (see [63] for a rigorous study).

Homeomorphisms of the background space alter landmarks and functions in a natural manner. This defines two group actions, which we choose to consider as left-actions in both cases:

**Definition 1** *Define the product  $gh$  of two diffeomorphisms by the (reversed) composition*

$$gh = h \circ g.$$

*Let  $g$  be a homeomorphism of  $\Omega$ . If  $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{L}^N(\Omega)$ , define*

$$g.\mathbf{x} = (g^{-1}(x_1), \dots, g^{-1}(x_N))$$

*and if  $I \in \mathcal{F}_d(\Omega)$ , define  $g.I$  by*

$$g.I(x) = I \circ g(x)$$

We shall return to these group actions later. For now on, we want to be able to quantify the size of a homeomorphism  $g$ , that is, to attribute to  $g$  a number  $E(g)$  which says how far  $G$  is from the identity. We start

reviewing a number of methods for this purpose. The finality of this study is, of course, to enable the comparison between landmarks and functions by evaluating how much distorsion is required to match two sets of landmarks or two functions.

Any diffeomorphism  $g$  can be represented as  $g = \text{id} + u$  where  $\text{id}$  is the identity. The reverse of course is false: if  $u$  is given,  $\text{id} + u$  is not necessarily a diffeomorphism, unless  $u$  is small. This is important when the matching problem is formulated by a variational formula in  $u$ : the best  $u$  (according to the variational formulation) may induce a non-bijective deformation (with, for example, foldings).

## 2.2 Standard functional norms

To measure the size of  $g = \text{id} + u$ , it is natural to quantify how far  $u$  is from zero by using one of the numerous norms which can be derived from functional analysis. It is clearly out of the scope of this tutorial to provide any kind of panorama of available norms from this point of view. However, from the computer-vision litterature, it seems that essentially two kinds of approaches have been followed to define such norms: the first one is based on a control of the derivatives of  $u$ , yielding norms similar to Sobolev's norms, and the second one controls expansions of  $u$  over a suitably chosen orthogonal basis of the space of functions defined on  $\Omega$  with value in  $\mathbb{R}^d$ .

For example, in one dimension, many authors simply measure the size of a diffeomorphism by the  $\mathcal{L}^2$  norm of its differential

$$E(g) = \int_{\Omega} (\dot{g}(x) - 1)^2 dx .$$

More generally,  $E(g)$  can be something of the kind (in any dimension)

$$E(g) = \|u\|^2 = \int_{\Omega} |Lu|^2 dx .$$

where  $L$  is a linear differential operator of dimension  $d$  and of degree  $m$ , in other terms,  $|Lg|^2$  is a sum of squared partial derivatives of components of  $g$ , of order at most  $m$ . There are several theoretical and practical reasons, related to Sobolev's inclusion theorem, to advise that the order  $m$  should increase with the dimension  $d$ .

To cite a few examples among many, the classical Horn and Schunck approach for motion estimation simply uses the  $\mathcal{L}^2$  norm of  $u$ , so that  $m = 0$ . Regularization of order 1 in 1D is the standard choice in curve matching. For landmark matching with splines,  $m$  is typically 2 in dimensions 2 or 3.

Another way to express the norm of a function uses the coefficients of its expansion over a suitably chosen orthonormal basis of  $\mathcal{L}^2(\mathbb{R}^k)$ . If  $f_1, \dots, f_n, \dots$  is such a basis, and  $v$  is defined on  $\Omega$ , with values in  $\mathbb{R}$ , the  $k$ th coefficient of  $v$  is

$$v_k = \int_{\Omega} v(x) f_k(x) dx$$

and, fixing a sequence of positive numbers  $c_1, \dots, c_k, \dots$ , we may set

$$\|v\| = \sum_{k=1}^{\infty} c_k v_k^2$$

This converges for all  $v$  provided the sequence  $(c_k)$  is bounded, but interesting sequences are in fact such that  $c_k$  tends to infinity: in this case, for a function  $v$ , the constraint that  $\|v\| < \infty$  generally corresponds to a smoothness condition (because functions  $f_k$  for large  $k$  generally carry high frequency components of  $v$ , for more details, see [46]). The advantage of such a decomposition is that it generally requires a smaller number of parameters to represent a function, and therefore leads to faster numerical implementations. Moreover, because the functions  $f_k$  analyse higher frequencies when  $k$  increases, this formulation leads naturally to multi-resolution algorithms. The drawback is that the evaluation of the function  $v$  (ie, the computation of  $v(x)$  for a given  $x$ ) become costly, which can have some importance for applications.

Both formulations are equivalent when the basis  $f_k$  diagonalizes the squared operator  $L^*L$  where  $L^*$  is defined by, for any  $u$  and  $v$ ,  $C^\infty$  with compact support in  $\Omega$

$$\int_{\Omega} u.Lv dx = \int_{\Omega} v.L^*u dx$$

Assuming that  $L^*L$  can be diagonalized in an orthonormal basis,<sup>1</sup> then, if  $L^*L f_k = c_k f_k$  and  $(f_k)$  is orthonormal, we have

$$\int_{\Omega} |Lv|^2 = \sum_{k=1}^{\infty} c_k v_k^2$$

This representation has been used to sample and estimate deformable templates in [4, 22, 1], [63].

---

<sup>1</sup>For example if  $L$  is an elliptic differential operator

### 2.3 Hyperelastic models

Elasticity theory starts with a series of universal principles, like material indifference, isotropy, to provide ways to quantify the amount of deformation exercised on an object. From these principles (which retain some validity for visual object models), and the assumptions that the elastic energy only depends on the deformation gradient, one introduces the notion of *hyperelastic materials*: if  $g$  is, as above, a diffeomorphism defined on some open set  $\Omega$  in  $\mathbb{R}^3$ , and  $Dg$  is the derivative of  $g$  (a  $3 \times 3$  matrix), the hyperelastic assumptions states that the energy should have the form

$$E(g) = \int_{\Omega} W(x, C) dx$$

where  $C = {}^t Dg.Dg$  is the Cauchy-Green tensor. If the material is homogeneous,  $W$  does not depend on  $x$ . For homogeneous, isotropic materials,  $W$  only depends on the eigenvalues of  $C$ . In this last case, it is just only necessary to define a function of two variables in 2D (three in 3D) to provides  $W(C)$  in function of the 2 (or 3) eigenvalues of  $C$ .

To define such functions, many authors simply have borrowed standard elasticity models, although one may argue about going so far in considering an image, for example, as a rubber plate. The simplest non-linear model is the Saint Venant-Kirchhoff material for which (letting  $E = (C - I)/2$ )

$$W(C) = \frac{\lambda}{2} [\text{trace}(E)]^2 + \mu \text{trace}(E^2)$$

$\lambda$  and  $\mu$  are the *Lamé* coefficients. These are used, for example, in [54].

Even more often, the linearized version of the model is used: if  $g = \text{id} + u$ , and  $u$  is small, the approximation  $E \simeq (Du + {}^t Du)/2$  is valid. Linear elasticity consists in replacing  $E$  by this approximation, yielding an energy which is quadratic in  $u$ , and linear Euler equations.

Other kinds of energies have been proposed in [34], requiring, in particular that  $W(C)$  tends to infinity when  $C$  tends to infinity, but also when  $\det(C)$  tends to 0, which is not the case for Saint Venant-Kirchhoff materials. For their application (face comparison), they used

$$W(C) = W(\alpha, \beta) = (\alpha^2 + \beta^2)(1 + (\alpha\beta)^{-2}) \quad (1)$$

where  $\alpha$  and  $\beta$  are the eigenvalues of  $C$ .

## 2.4 Iterating small deformations: geodesic energies

In this section we define the cost of a deformation  $h$  as a sum of costs of elementary, infinitesimally small, deformations, the composition of which provides  $h$ .

Consider, for this purpose, the following situation: let  $v_1, \dots, v_n$  be small displacements on  $\Omega$ , and let  $\phi_i = \text{id} + v_i$ . Incrementally build a deformation by applying first  $\phi_1$ , then  $\phi_2$  and so on. The total deformation after  $k$  steps is

$$g_k = (\text{id} + v_k) \circ \dots \circ (\text{id} + v_1)$$

which yields

$$g_k - g_{k-1} = v_k \circ g_{k-1} \tag{2}$$

to which we add the constrain (boundary condition) that  $g_n = h$ .

Assume that a cost  $\|v\|$  can be associated to any infinitesimal displacement  $v$ , through a functional norm on  $v$  (such as the ones previously described). Define the energy of the deformation process simply as the sum

$$\Gamma(v_1, \dots, v_n) = \|v_1\|^2 + \dots + \|v_n\|^2$$

Now let  $n$  tend to infinity, and exchange the discrete indices  $k$  with a continuous variable  $t \sim k/n$  in  $[0, 1]$ . Introducing the time-indexed displacements  $v(t, \cdot)$  (which are seen as the limit of  $(v_{[nt]})/n$  for large  $n$ ), and letting  $g(t, \cdot)$  be the limits of  $g_{[nt]}(\cdot)$ , equation (2) becomes

$$\frac{dg}{dt}(t, \cdot) = v(t, g(t, \cdot)) \tag{3}$$

the boundary conditions being now  $g(0, \cdot) = \text{id}$  and  $g(1, \cdot) = h$ . The cost becomes

$$\Gamma(v) = \int_0^1 \|v(t)\|^2 dt$$

(recall that  $v(t)$  itself is a function defined on  $\Omega$  and that the norm is functional).

This yields a new formulation for evaluating the energy of  $h$ : let  $\mathcal{V}_h$  be the set of all time-dependent displacements  $t \mapsto v(t, \cdot)$  in  $\Omega$  such that the solution of the ordinary differential equation

$$\frac{dy}{dt} = v(t, y)$$

with initial condition  $y(0) = x$  satisfies  $y(1) = h(x)$ . We let

$$E(h) = \inf_{v \in \mathcal{V}_h} \int_0^1 \|v(t)\|^2 dt$$

At first sight, this seems quite an awkward way to define a cost function associated to  $h$ . In particular, introducing a new variational problem seems quite inefficient. Moreover, it is not obvious that the set  $\mathcal{V}_h$  has any element at all, and even less obvious is the problem of characterizing its elements.

However, this point of view has several advantages. First the difficulty of characterizing the set  $\mathcal{V}_h$  can be overcome by working directly with  $v$  instead of  $h$ . Indeed, in practice,  $h$  is not given *per se*, but rather as part of a problem in which one must find *the smallest*  $h$  subject to some constraints. It suffices to replace the problem by: finding the smallest  $v$  such that the solution at time 1 of the previous ordinary differential equation satisfies the same constraints. The consequence of this is to add an extra dimension to the problem, but this is counter-balanced by the simplicity of the cost function when expressed in terms of  $v$ . Also, the continuous time deformation process which is obtained when integrating  $v$  can also be of some interest, for motion interpolation, or to generate morphings in computer graphics.

Another advantage is that the obtained cost function is closely related to a notion of *distances* between diffeomorphisms, which would in turn provide distances between the compared objects (see section 5.4). Triangular inequality and symmetry is much harder to obtain with an elasticity theory formulation.

A last advantage in formulating the problems in terms of  $v$  is that  $v$  naturally belongs to a linear space. Linear combination of diffeomorphisms are not necessarily diffeomorphisms, but linear combinations of time dependent displacements remain time dependent displacements. This allows to consider using methods of linear analysis, or gaussian statistical modeling without harming the mathematical rigor.

## 2.5 Learning energies from data-sets

### 2.5.1 Generalities

We now pass to methods for energy definition based on a statistical analysis of a data set composed with observed deformations. Although the presentation is quite different, this is very close to the notion of principal warps described in [15]. Principal component analysis has also been used in [20] (among other works) to quantify variations of landmarks configurations (but without considering diffeomorphisms, as we do here).

We start with an abstract description of principal component analysis, which is more appropriate to handle infinite dimensions.

Let  $u_1, \dots, u_N$  be elements of a vector space  $H$ , equipped with an inner product  $\langle \cdot, \cdot \rangle$ . One can see principal component analysis as a way to represent the  $u_k$  under the form

$$u_k = \bar{u} + \sum_{i=1}^p \alpha_{ki} e_i + R_k$$

with  $\bar{u} \in H$ ,  $e_1, \dots, e_p$  being an orthonormal family and  $R_k$  being as small as possible, more precisely minimizing the mean-square error

$$S = \frac{1}{N} \sum_{k=1}^N \|R_k\|^2$$

When  $\bar{u}$  and the  $e_i$  are given,  $\sum_{i=1}^p \alpha_{ki} e_i$  must be the orthogonal projection of  $u_k - \bar{u}$  on the space spanned by  $(e_1, \dots, e_p)$ , which implies that  $\alpha_{ki} = \langle u_k - \bar{u}, e_i \rangle$ . Still fixing the  $e_i$ , it is clear that one should set  $\bar{u} = \frac{1}{N} \sum_{k=1}^N u_k$ . In the rest of the presentation, we assume that  $\bar{u} = 0$  (or, more precisely, we replace  $u_k$  by  $u_k - \bar{u}$  without changing the notation). Thus, the family  $(e_k)$  must be selected to minimize

$$\Sigma = \frac{1}{N} \sum_{k=1}^N \left\| u_k - \sum_{i=1}^p \langle u_k, e_i \rangle e_i \right\|^2$$

which writes

$$\Sigma = \frac{1}{N} \sum_{k=1}^N \|u_k\|^2 - \sum_{i=1}^p \sum_{k=1}^N \langle u_k, e_i \rangle^2$$

Introduce, on  $H$ , a new inner-product,  $\langle \cdot, \cdot \rangle_u$ , given by

$$\langle x, y \rangle_u = \frac{1}{N} \sum_{k=1}^N \langle u_k, x \rangle \langle u_k, y \rangle$$

let  $\| \cdot \|_u$  be the associated (semi-)norm:  $(e_1, \dots, e_p)$  must therefore maximize  $\sum_{i=1}^p \|e_i\|_u^2$  and form an orthonormal family for the initial inner-product of  $H$ .

The quadratic form  $\|\cdot\|_u^2$  can be diagonalized in an orthonormal basis  $(f_n, n \geq 0)$  of  $H^2$ . Letting  $\lambda_n = \|f_n\|_u$ , and assuming that they are ordered in decreasing order, it is a standard fact that the optimal choice is to let  $e_i = f_i$  for  $i = 1$  to  $p$ .

### 2.5.2 Implementation

In finite dimension, the points  $u_1, \dots, u_N$  can be considered as vectors expanded over a given orthonormal basis of  $H$ . If the selected inner-product on  $H$  is associated to a definite positive symmetric matrix  $A$  (ie.  $\langle x, y \rangle = {}^t x A y$ ), the semi-norm  $\|\cdot\|_u$  is associated to the matrix

$$Q_u = \frac{1}{N} A \cdot \left( \sum_{k=1}^N u_k {}^t u_k \right) \cdot A,$$

and the problem is to find a basis orthogonal for  $Q_u$  and orthonormal for  $A$ . This can be obtained by diagonalizing the product

$$A \left[ \frac{1}{N} \sum_{k=1}^N u_k {}^t u_k \right]$$

yielding eigenvectors  $f'_i$ , and eigenvalues  $\lambda_i$ , and then setting  $e_i = A^{-1} f'_i$ .

In infinite dimension (which is the case of interest here, since we are dealing with deformations), there are two options. The first one is to choose from the beginning a finite dimensional representation, by projection on an *a priori* finite dimensional basis (Fourier, wavelets, splines, ...), and apply the previous formulas. Another point of view is to look for  $e_i$  under the form

$$e_i = \sum_{k=1}^N \lambda_{ik} u_k.$$

since it is clear that the optimal basis belongs to the vector space spanned by  $(u_1, \dots, u_N)$ .

The solution is then to look for a common orthogonal basis for the  $N \times N$  matrices  $A$  and  $B$ , with  $a_{ij} = \langle u_i, u_j \rangle$  et  $b_{ij} = \langle u_i, u_j \rangle_u$ , which is also a standard problem in matrix computation.

---

<sup>2</sup>The problem is in fact finite dimensional, since  $\|\cdot\|_u$  vanishes for any vector orthogonal to the vector space spanned by  $u_1, \dots, u_k$

When the  $u_i$  are functions, there are many choices for the inner-product on  $H$ . The  $\mathcal{L}^2$ -product is the most standard, but there may be some interest in selecting a product which controls more derivatives, for example

$$\langle u, u \rangle = \int_{\Omega} |\Delta u + \alpha u|^2 dx.$$

### 2.5.3 Learning deformation energies

Principal component analysis is widely applied in image analysis. Cootes et al. ([20]) have introduced this technique for shape detection and comparison, by decomposing sets of landmarks over “eigenshapes”. It is also widely used for face recognition, or facial expression synthesis ([14, 67, 35, 68]). Let us rapidly describe how to use it to quantify deformations.

Consider a collection of images  $(I_1, \dots, I_N)$  and a template  $I_0$ , and assume that it has been possible to estimate a family of deformations  $u_1, \dots, u_N$  such that, for all  $k$ ,  $I_k \simeq I_0 \circ (\text{id} + u_k)$ , for example by the landmark-spline interpolation which will be described in the next section: this will form the training set.

For any given inner-product on the space of deformations, it is then possible to compute an orthonormal basis of deformations  $e_1, \dots, e_p$ , which is orthogonal for  $\langle \cdot, \cdot \rangle_u$ , taking eigenvalues  $\lambda_i = \langle e_i, e_i \rangle_u$  in descending order. One consistent choice for the size of a deformation  $u$  can then be

$$E(u) = \sum_{i=1}^p \frac{\langle u, e_i \rangle^2}{\lambda_i}.$$

## 3 Landmark-based matching

The issue in landmark-based matching, is to interpolate a discrete matching between a finite set of landmarks in  $\Omega$  to obtain a dense diffeomorphism in  $\Omega$ . In this sense, the associated techniques and solutions are closely related to spline fitting.

So, let  $\Omega$  be our background space, and  $(x_1, \dots, x_N), (y_1, \dots, y_N)$ , two sets of  $N$  matched landmarks in  $\Omega$ . The problem is to find a diffeomorphism  $g$  of  $\Omega$ , with minimal size, such that, for all  $i$ ,  $g(x_i) = y_i$  (*exact matching*), or, for all  $i$ ,  $g(x_i) \simeq y_i$  (*inexact matching*), the latter case being generally expressed as an error penalty incorporated into a variational setting.

Before exploring with more details the landmark matching methods, we start with a few remarks on spline theory.

### 3.1 Splines

In an abstract setting, spline fitting can be considered a particular cases of the following situation. We let  $\mathcal{H}$  be a Hilbert space, and let  $f_1, \dots, f_N \in \mathcal{H}$ , and  $c_1, \dots, c_N \in \mathbb{R}$ . We can formulate two problems:

1. Minimize  $\|h\|$  over  $\mathcal{H}$  subject to the constraints  $\langle f_i, h \rangle = c_i$  for  $i = 1, \dots, N$ .
2. Fixing  $\lambda > 0$ , minimize  $\|h\|^2 + \lambda \sum_{i=1}^N (\langle f_i, h \rangle - c_i)^2$

The first problem corresponds to interpolation, or exact matching, the second one to smoothing, or approximate matching, and both are solved by elementary linear algebra. It is indeed clear that, in both cases, the constraints are not affected by if  $h$  is replaced by  $h + v$  where  $v$  is orthogonal to all the  $f_i$ , so that the solution must in fact be searched in the linear space spanned by  $f_1, \dots, f_N$ .

So, introduce the  $N \times N$  matrix  $S$  with  $S_{ij} = \langle f_i, f_j \rangle$ , and express the unknown  $h$  as a linear combination

$$h = \sum_{i=1}^N \alpha_i f_i$$

Problem 1 is to minimize  ${}^t\alpha S\alpha$  subject to the constraint  $S\alpha = c$  (where  $\alpha$  and  $c$  are vectors with components  $\alpha_i$  and  $c_i$  respectively), and problem 2 is to minimize

$${}^t\alpha S\alpha + \lambda {}^t(S\alpha - c)(S\alpha - c)$$

Assume, to simplify, that  $S$  is invertible, so that no linear constraint can be deduced one others. The solution of problem 1 is in fact directly obtained from the constraint: it is  $\alpha = S^{-1}.c$ . For the second, it is  $\alpha = \lambda(I + \lambda S)^{-1}.c$  (for which the invertibility of  $S$  is in fact not required).

The relation with splines is as follows: spline interpolation corresponds to finding a real-valued function  $h$  (defined on  $\Omega$ ), as smooth as possible, such that  $h(x_i) = c_i$ , with given  $x_i \in \Omega$  and  $c_i \in \mathbb{R}$ . The smoothness of  $h$  is evaluated through a norm of the kind

$$\|h\|_L = \int_{\Omega} |Lh|^2 dx$$

where  $L$  is, say, a differential operator. This norm defines a Hilbert space of functions  $\mathcal{H}_L$ , with the inner-product

$$\langle h, g \rangle_L = \int_{\Omega} LhLg dx.$$

The constraints  $h(x_i) = c_i$  are linear in  $h$ , and the issue, to fit in the abstract setting, is whether there exists an element  $f_{x_i}$  in  $\mathcal{H}_L$  such that, for all  $h \in \mathcal{H}_L$

$$h(x_i) = \langle f_{x_i}, h \rangle_L$$

If this can be done<sup>3</sup>, the solution of the interpolation problem is given by a linear combination of the  $f_{x_i}$ , which coefficients are simply obtained by applying the inverse of the matrix of inner-products of the  $f_{x_i}$  to the values of the constraints  $c_i$ . A similar conclusion can be drawn if we replace this exact interpolation problem by an inexact form, which consists in minimizing

$$\|h\|_L + \lambda \sum_{i=1}^N (h(x_i) - c_i)^2$$

It must be noted that the inner-products  $\langle f_{x_i}, f_{x_j} \rangle$  are, by construction, given by  $f_{x_i}(x_j)$  ( $f$  is self-reproducing), so that their computation is immediate.

So everything boils down to the existence of the  $f_x$ . The theoretical arguments for proving this existence are linked to Sobolev's inclusion theorem, but, for practical purposes, it is also necessary to know an analytical expression for them.

Introduce the dual operator  $L^*$ , such that, for all  $g$  and  $h$  with compact support in  $\Omega$ ,

$$\int_{\Omega} (Lh)g = \int_{\Omega} (L^*g)h$$

and let  $K = L^*.L$ . Since, for all  $x$ ,

$$h(x) = \langle f_x, h \rangle_L = \int_{\Omega} Lf_x Lh dy = \int_{\Omega} f_x K h dy$$

the function  $(x, y) \mapsto f_x(y)$  is, by definition, the Green kernel of the operator  $K$ . So, the practical applicability of the spline interpolation in this setting depends on whether the Green kernel of  $K$  is known or not, provided of course it exists at all (there are other complications linked to boundary conditions, which affect the Green kernel, but we here omit them).

A well-know case in which explicit expressions for Green functions are available is when  $L$  is a variant of the Laplacian (with simple enough boundary conditions). To get more generality, it is possible to start directly with a function

$$(x, y) \mapsto f_x(y)$$

---

<sup>3</sup>This is equivalent, by the Riesz representation theorem, to the continuity of the evaluation mapping  $h \mapsto h(x)$  for the norm  $\|\cdot\|_L$

since everything is in fact depending on it. Under some conditions, this function is associated to some operator  $L$  (not necessarily differential), which provides a variational basis to the interpolation method. A sufficient set of conditions can be

- $f$  is symmetric:  $f_x(y) = f_y(x)$
- $f$  is squared-integrable: for all  $x$

$$\int_{\Omega} (f_x(y))^2 dx < \infty$$

- $f$  is continuous
- $f$  induces a positive operator on  $\mathcal{L}^2$ : for all  $u \in \mathcal{L}^2$

$$\int_{\Omega} u(x)u(y)f_x(y)dx dy$$

The last requirement is realized if  $f_x(y) = F(x - y)$  where  $F$  is the fourier transform of a positive, even function.

In [6], [5], the requirements are further specialized so as to take for  $f$  a radial basis function:  $f_x(y) = G(|x - y|)$ , the simplest example being the Gaussian

$$G(t) = \exp\left(-\frac{t^2}{\sigma^2}\right)$$

### 3.2 Bookstein's splines

Landmark-based matching differs from the previous context by the fact that the unknown function is vector valued. In fact, if  $x_1, \dots, x_N$  and  $y_1, \dots, y_N$  are two matched sets of landmarks, one should find a diffeomorphism  $h : \Omega \rightarrow \Omega$  such that  $h(x_i) = y_i$  (equivalently, one searches a displacement  $u = h - \text{id}$  such that  $u(x_i) = y_i - x_i$ ).

Bookstein (see [15]) proposes to apply spline interpolation to each component of  $u$ . This is the simplest approach, and we will keep to this setting. We now describe how this is implemented. Let  $U(r) = r^2 \log r$ : this function is such that, for any smooth function  $f$ , one has

$$f(x) = \int_{\mathbb{R}^2} U(|x - y|) \Delta^2 f(y) dy$$

where  $\Delta^2$  is the iterated Laplacian. To retrieve the previous setting, one must consider the inner-product

$$\langle f, g \rangle = \int_{\mathbb{R}^2} \Delta f \Delta g dx. \quad (4)$$

which is such that  $\langle f_x, g \rangle = g(x)$  where  $f_x(y) = U(|x - y|)$ . So let  $c_i$  be one of the components of  $y_i - x_i$ , so that the constraint  $h(x_i) = c_i$  writes  $\langle h, f_{x_i} \rangle = c_i$ . We have

$$S_{ij} = \langle f_{x_i}, f_{x_j} \rangle = U(|x_i - x_j|)$$

The direct application of the previous section would yield that the interpolation,  $\gamma$ , is given by

$$z(x) = \sum_{i=1}^N \alpha_i U(|x - x_i|)$$

with  $\alpha = S^{-1}c$  in the case of exact matching, and  $\alpha = \lambda(I + \lambda S)^{-1}.c$  for inexact matching. However, the situation is slightly more complicated, because the inner-product (4) is degenerate:  $\langle f, f \rangle = 0$  when  $f$  is linear.<sup>4</sup> To take this into account, the procedure must be slightly modified: since an affine displacement is invisible for the inner product, it is “free” to replace the assignments  $c_i$ , for  $i = 1, \dots, N$  by  $c_i - {}^t a x_i - b$ , for any given  $a \in \mathbb{R}^2$  and  $b \in \mathbb{R}$ . Thus, the interpolation problem becomes: minimize

$${}^t \alpha S \alpha$$

with the constraint  $S\alpha + Q\gamma = c$  where  $\gamma = {}^t(a_1 a_2 b)$  is a  $3 \times 1$  matrix and  $Q$  is a  $N \times 3$  matrix, given by, letting  $x_i = (x_i^1, x_i^2)$ :

$$Q = \begin{pmatrix} x_1^1 & x_1^2 & 1 \\ \vdots & \vdots & \vdots \\ x_N^1 & x_N^2 & 1 \end{pmatrix}$$

solving this problem in  $(\alpha, \gamma)$  yields

$$\hat{\gamma} = ({}^t Q S^{-1} Q)^{-1} {}^t Q c$$

---

<sup>4</sup>One should be more precise (cf [45]): the Hilbert space  $\mathcal{H}$  under consideration is the Beppo-Levi space composed with all (generalized) functions with square integrable second derivatives. Among these functions, it can be shown that  $\langle f, f \rangle = 0$  if and only if  $f$  is linear.

and  $\hat{\alpha} = S^{-1}(c - Q\hat{\gamma})$ .

The smoothing problems requires minimizing

$${}^t\alpha S\alpha + \lambda^t(S\alpha + {}^t\alpha x_i + b - c)(S\alpha + {}^t\alpha x_i + b - c)$$

and its solution is formally similar to the previous one, simply replacing  $S$  by  $S_\lambda = S + (1/\lambda)I$  in the formulas.

When this is applied to both components of  $y_i - x_i$ , one obtains an function  $u$  such that  $u(x_i) = y_i - x_i$  for exact matching, which thus provides a smooth interpolation of the landmark matching. However, there is no constraint in this approach, which ensures that  $h(x) = x + u(x)$  is one-to-one: folding is indeed possible. An example of this kind of interpolation is provided in figure 2, together with an interpolation with a radial kernel ([6]).

### 3.3 Joshi's interpolation

To guarantee diffeomorphisms, Joshi has applied the approach of section 2.4 to the landmark matching framework. Returning to this formalism leads to introduce time-dependent displacements  $t \mapsto v(t, \cdot)$  in  $\Omega$ , and the ordinary differential equation

$$\frac{dy}{dt} = v(t, y)$$

Denote by  $h_v(t, x)$  the solution of this equation at time  $t$  with initial condition  $h_v(0, x) = x$ . The exact landmark matching problem, as formulated in [39] is to find  $v$  which minimizes

$$E(v) = \int_0^1 \int_\Omega |Lv(t)|^2 dx dt$$

subject to the constraint  $h_v(1, x_i) = y_i$  for each pair  $(x_i, y_i)$  of matched landmarks.

Denote by  $u_i(t) = h_v(t, x_i) - x_i$ , the trajectories of the displacements: given these trajectories, the optimal  $v$  is, at each time  $t$ , an interpolating spline of the previous kind, with the constraints  $v(t, x_i + u_i(t)) = \dot{u}_i(t)$  where the dot refers to time derivative. If  $(x, y) \mapsto f_x(y) = f(x, y)$  is the Green function associated to the problem, let  $S(t)$  be the matrix composed with the  $s_{ij} = f(x_i + u_i(t), x_j + u_j(t))$ . Let  $u_i(t) = (p_i(t), q_i(t)) \in \mathbb{R}^2$ , and denote by  $(v_1, v_2)$  the components of  $v$ . One has,

$$v_1(t, x) = \sum_{i=1}^N \alpha_i(t) f_{x_i + u_i(t)}(x)$$



Figure 2: Spline interpolation of landmarks. First row: original images with landmarks; second row: thin-plate (bookstein) interpolation (from the first image to the second and vice-versa) ; third row: interpolation using  $f_x(y) = \log(|x - y|^2 + c)$  (images taken from the Olivetti Face Database).

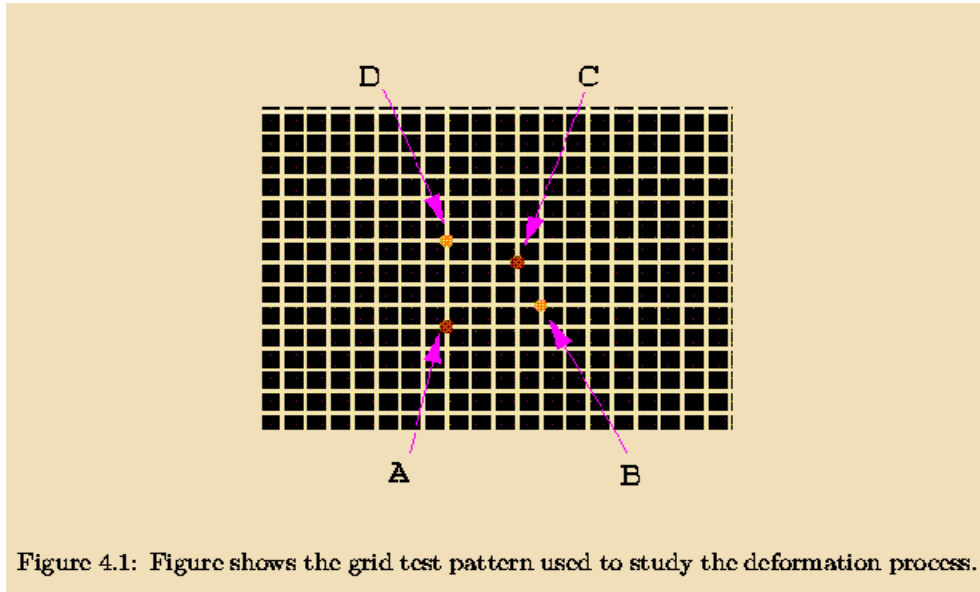


Figure 4.1: Figure shows the grid test pattern used to study the deformation process.

Figure 3: Figure taken from [39]: point A will be matched to B and C to D while corners are left unchanged

with  $\alpha(t) = S(t)^{-1}\dot{p}(t)$  and

$$v_2(t, x) = \sum_{i=1}^N \beta_i(t) f_{x_i+u_i(t)}(x)$$

with  $\beta(t) = S(t)^{-1}\dot{q}(t)$ . and the integrated energy is

$$E(v) = \int_0^1 {}^t\dot{p}(t)S(t)^{-1}\dot{p}(t)dt + \int_0^1 {}^t\dot{q}(t)S(t)^{-1}\dot{q}(t)dt$$

which is only expressed as a function of the  $N$  dimensional trajectories  $p$  and  $q$ , which are subject to the constraints  $(p(0), q(0)) = 0$  and  $(p(1), q(1)) = y - x$ . This can be minimized by gradient descent, and we refer to [39] for details. Figures 3 and 4 show an example of large deformation in which the introduction of the time variable yielded a significant improvement.

## 4 Dense matching

When landmarks are not available, or do not provide enough information to ensure accurate matching, dense matching methods can be used, pro-

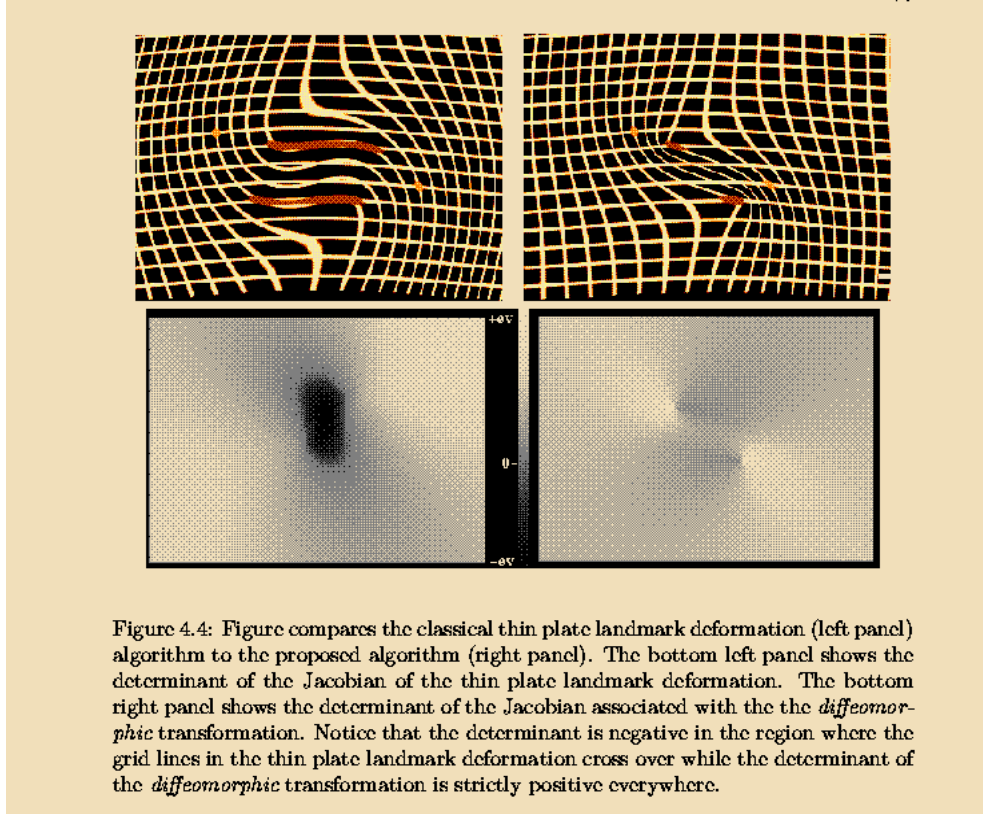


Figure 4.4: Figure compares the classical thin plate landmark deformation (left panel) algorithm to the proposed algorithm (right panel). The bottom left panel shows the determinant of the Jacobian of the thin plate landmark deformation. The bottom right panel shows the determinant of the Jacobian associated with the the *diffeomorphic* transformation. Notice that the determinant is negative in the region where the grid lines in the thin plate landmark deformation cross over while the determinant of the *diffeomorphic* transformation is strictly positive everywhere.

Figure 4: Figure taken from [39]

vided it is known that some quantity measured over the background space is (approximately) conserved through the matching.

Let, as before,  $\Omega$  be the background space, subject to deformations. Let the conserved quantities be functions  $I$  defined on  $\Omega$ , possibly assuming vector values. Let  $I$  and  $I'$  be such functions: the issue is to find a diffeomorphism  $g$  such that  $I \circ g$  is close to  $I'$  and  $g$  is as small as possible for some measure of size.

## 4.1 Variational approaches

Many algorithms start from a variational problem of this kind: find a diffeomorphism  $g$  which minimizes

$$U(g) = S(g) + \lambda D(I \circ g, I')$$

where  $\lambda$  is a parameter,  $S(g)$  is one of the measures of size which have been discussed in section 2, and  $D$  is a metric which measures the discrepancy between  $I \circ g$  and  $I'$ , thus after matching. The simplest choice for  $D$  is the  $\mathcal{L}^2$  norm

$$D(I \circ g, I') = \int_{\Omega} |I \circ g - I'|^2 dx$$

We now review a few examples from the literature.

### 4.1.1 1D Matching

One of the earliest problems in which warping solutions have emerged comes from speech recognition: given two acoustic signals recorded during a period of time by two different speakers the issue is to decide of their similarity without being affected by the natural distortions and variations which can be observed after recording a given spoken word. One of these distortions are the variations of instantaneous speech speed, which can be modeled as a diffeomorphism acting on the signals, as we did before, a signal  $I$  becoming  $I \circ g$ . For the first time in [55], an algorithm for “dynamic time warping” have been devised. In a discrete settings, for two sampled signals  $I = (I_1, \dots, I_N)$  and  $I' = (I'_1, \dots, I'_N)$ , the algorithms minimizes a functional of the kind

$$\Delta(I, I') = \sum_{k=1}^N d(I_k, I'_{\psi(k)})$$

where  $\psi(k)$  is the index for the homologous point of  $I_k$ . The minimization is subject to the constraint that  $\psi$  is increasing, and, typically, a boundedness constraint on the variations  $\psi(k+1) - \psi(k)$ .

In [53], 1D speech signals  $I$  and  $I'$  are matched by minimizing

$$U(g) = \arccos \left[ \int_0^1 \sqrt{\dot{g}_x} dx \right] + \lambda \int_0^1 (I(g(x)) - I'(x))^2 dx$$

This energy arises in fact from the geodesic design of section 2.4 (we shall return on it later).

In computer vision, 1D-matching is applied to track and compare contours or shapes.

Contours can be modeled as planar (or 3D) curves, ie. functions defined on an interval with values in  $\mathbb{R}^2$  (or  $\mathbb{R}^3$ ). Their comparison can be directly based on matching points as they are located in space, for example by minimizing

$$\Delta(I, I') = E(g) + \lambda \int_0^1 |I'(t) - I \circ g(t)|^2 dt$$

However, most of the applications imply comparing shapes, which are contours seen up to translations, rotation and sometimes scaling. A natural point of view, then, is to match invariant representations of the curves, like, for example, the curvature function (invariant to rotation and translation), the tangent angle (invariant to translation and scaling), both expressed in function of arc-length, or higher order representations. For example, in [19], the following energy is proposed: let  $I$  and  $J$  be curves, indexed by arc-length, such that  $I$  has length 1 and  $J$  has length  $\alpha$ . the matching is searched as a one-to-one mapping  $h : [0, 1] \rightarrow [0, \alpha]$  by minimizing

$$U(h) = \int_0^1 (\kappa_I(s) - \kappa_J \circ h(s))^2 ds + \lambda \int_0^1 \left| \frac{d}{ds} (I(s) - J \circ h(s)) \right|^2 ds$$

where  $\kappa_I$  and  $\kappa_J$  are the curvatures of  $I$  and  $J$ . The energy can be rewritten in a different manner, introducing the unit tangents  $\tau_I$  and  $\tau_J$ , and expanding the second integral to get

$$1 + \int_0^1 \dot{h}_s^2 ds - 2 \int_0^1 \dot{h}_s \langle \tau_I, \tau_J \circ h \rangle(s) ds$$

If  $(s' \mapsto J(s'))$  is reparametrized by  $s' \mapsto s'/\alpha$  and  $g(s) = f(s)/\alpha$ , the formula above is

$$1 + \alpha^2 \int_0^1 \dot{g}_s^2 ds - 2\alpha \int_0^1 \dot{g}_s \langle \tau_I, \tau_J \circ g \rangle(s) ds$$

which is also

$$1 + \alpha^2 \int_0^1 \dot{g}_s^2 ds - 2\alpha \int_0^1 \langle \tau_I \circ g^{-1}, \tau_J \rangle(s) ds.$$

This finally yields

$$\begin{aligned} U(g) &= \lambda \alpha^2 \int_0^1 \dot{g}_s^2 ds + \int_0^1 (\kappa_I(s) - \kappa_J \circ g(s))^2 ds \\ &\quad + \lambda \left( 1 - 2\alpha \int_0^1 \langle \tau_I \circ g^{-1}, \tau_J \rangle(s) ds \right) \end{aligned}$$

which exhibits an elastic deformation term, and a comparison of both curvatures and tangent angles.

In [12], an axiomatic approach led to a series of energies to match plane curves on the basis of their curvatures; for example, one of them takes the form

$$U(g) = \int_0^1 |\dot{g} - 1| ds + \lambda \int_0^1 (\dot{g} + 1) |f(\kappa_I \circ g) - f(\kappa_J)| ds$$

where  $f(\kappa) = c\kappa - \text{sign}(\kappa)e^{-\alpha\kappa}$ , or

$$U(g) = \int_0^1 |\dot{g} - 1| ds + \lambda \int_0^1 |\dot{g}\kappa_I \circ g - \kappa_J| ds$$

In [71], matching was performed using

$$U(g) = \arccos \frac{\sqrt{2}}{2} \int_0^1 \sqrt{\dot{g} (1 + \langle \tau_I \circ g, \tau_J \rangle)} ds$$

To match 3D curves, the authors in [11] have used a characterization through the Frénet's frame instantaneous rotation, which is a 3 by 3 matrix depending in the curvature and the torsion of the curve. The final formula takes the form

$$U(g) = \int_0^1 |\dot{g} - 1| ds + \int_0^1 \|F_I - F_J \circ g\|^2 ds$$

where  $\|\cdot\|$  is the Hilbert-Schmidt norm  $\|A\|^2 = \text{trace}({}^t A.A)$ .

Most of the times, the energies so-defined can be very efficiently minimized by dynamic programming algorithms (which yielded the term *dynamic time warping* in the speech recognition literature). We shall not discuss dynamic programming issues here (see [64] for other details).

### 4.1.2 Image and volume matching

From a theoretical point of view, 3D matching is not very different from 2D problems (there are important peculiarities in 1D models). Obviously, the main issue in the change of dimensionality from 2D to 3D is the dramatic increase of the number of unknowns for numerical procedures.

#### Optical flow and stereo vision

2D matching has a long history in image analysis, since it is closely related to two essential issues, which are motion estimation and stereo-vision. Although these problems are rarely written with an elastic matching formulation, it is quite interesting to see how they fit in the general framework.

For motion estimation, a displacement field  $u$  has to be estimated between two consecutive images  $I$  and  $I'$  so that  $I(x + u) \simeq I'(x)$ . There is a huge literature dealing with optical flow, and shall make no attempt to give any account of it. However, a loose interpretation of the variational methods which have been used in this context, yields an energy which takes the form ([37]), letting  $g = \text{id} + u$ ,

$$U(g) = \int_{\Omega} \alpha^2 |u(x)|^2 dx + \int_{\Omega} |I(g(x)) - I'(x)|^2 dx \quad (5)$$

the last term being often linearized, assuming small  $u$ , as

$$I(x) - I'(x) + \nabla I(x) \cdot u(x)$$

which induces the “optical flow equation”.

In stereo vision, matching two images induces knowledge of a depth map in the visible scene recorded by two calibrated cameras. This case is more peculiar, because there a geometric constraint must be taken into account, which is that homologous points in images are known in advance to lie on prescribed lines called epipolar. After an image transformation, it can be assumed that the epipolar are horizontal lines, which implies that the displacement which must be estimated has only one non-vanishing coordinates, the first one. This fact apart, the same kind of techniques can be used as in motion estimation.

#### Face comparison

Another range of application in which elastic matching have been extensively applied is face comparison. There are mainly three factors of variations in

face images, for the same individual, namely illumination, pose, and facial expression. Only the last one is related to the topic of this talk, but we can refer to the recent book [34] to provide an extensive introduction to the global problem.

The most popular approach to the problem, eigenfaces, does not explicitly take into account this deformation component. This approach is essentially based on principal component analyses of face images, simply registered by an affine (rigid) transformation.

Elastic matching, however, is extensively used by C. Von der Malsburg school ([21, 69]), in which face comparison is performed with a discrete counterpart of an elastic deformation energy, the compared quantities being Gabor transforms of the images rather than the plain intensities. The matching process is interpreted as a neural network, and correlated to biological evidences.

In [34], the whole process of illumination variation, pose variation and rigid deformations is addressed by a single (and quite complex) matching energy. Among the important features of this energy, is the fact that the elastic deformations are controlled by a non-conventional term, given in equation (1), and that grey-level intensities and an edge map are simultaneously matched.

### Medical imaging

The widest range of application of elastic matching is probably provided by medical images, in 2D and 3D. New imaging devices provide images in different modalities, and of increasing quality, with a strong need of processing (construction of anatomic atlases, detection of variations in functional data, registration of images taken from different patients, and/or different modalities, help for surgery ... ). One of the most developed area is related to brain imaging, and [62] can serve as a beautiful reference for this topic.

### Viscous formulation

The geodesic deformation cost of section 2.4 is used in [63]. The problem is to minimize

$$U(v) = \int_0^1 \int_{\Omega} |Lv|^2 dx dt + \lambda \int_{\Omega} (I \circ g_v(1, x) - I'(x))^2 dx$$

in  $v$ , where, for all  $x \in \Omega$ ,  $g_v(t, x)$  is the value at time  $t$  of the solution of the ODE  $\dot{y} = v(t, y)$  with initial value  $y(0) = x$ . Since this computation is

somewhat tricky, we here give the expression of the gradient (for the  $\mathcal{L}^2$ -norm, see remarks below), and thus formally compute the variation  $U(v + h) - U(v)$  for small  $h$ .

We thus consider the issue of minimizing

The first term in  $U$  is

$$\int_0^1 \int_{\Omega} |Lv|^2 dx dt$$

This is a quadratic term in  $v$ . Once discretized, its differential can be very easily computed, which is very important from a numerical point of view. The analytical expression involves the dual operator  $L^*$ , defined by, for all  $v, w$ :<sup>5</sup>

$$\int_0^1 \int_{\Omega} \langle Lv, w \rangle dx dt = \int_0^1 \int_{\Omega} \langle v, L^*w \rangle dx dt$$

so that, at first order,

$$\int_0^1 \int_{\Omega} |L(v + h)|^2 dx dt \simeq \int_0^1 \int_{\Omega} |Lv|^2 dx dt + 2 \int_0^1 \int_{\Omega} \langle L^*Lv, h \rangle dx dt.$$

The differential of the second term is harder to compute, and we shall not detail here the computation which has been led in [13], and give only the results which require some additional notation. Let  $h_v(t, \cdot)$  be the inverse of  $g_v(t, \cdot)$ , and  $j_v(t, x) = g_v(1, h_v(t, x))$ . Let  $S(t, y) = I(t, h_v(t, y))$  et  $S'(t, y) = I'(t, j_v(t, y))$ . Let

$$q(v) = \int_{\Omega} (I \circ g_v(1, x) - I'(x))^2 dx.$$

We have

$$q(v + h) = q(v) + \int_0^1 \int_{\Omega} \langle b_v(t, y), h(t, y) \rangle dt dy + o(h)$$

with

$$b_v(t, y) = -2 \det(Dh_v(t, y))^t DS'(t, y) (S(t, y) - S'(t, y))$$

with the notation  $DF(t, y)$  for the matrix composed with the spatial derivatives of a differentiable function  $F$ . Thus, the  $\mathcal{L}^2$ -gradient of  $U$  is

$$2L^*Lv - 2\lambda \det(Dh_v(t, y))^t DS'(t, y) (S(t, y) - S'(t, y))$$

Figure 5 provides an example of implementation of gradient-descent using this energy for shape matching.

---

<sup>5</sup>We assume null boundary conditions on  $\Omega$

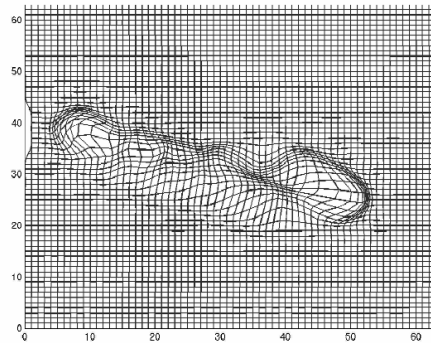


Figure 5: Matching upper right hippo-campus shape to upper left. Lower image: deformation applied to a grid

### 4.1.3 Numerical implementation

#### Discretization

The first remark concerns the representation of the unknown function  $g$  (or  $u = \text{id} - g$ , or  $v$  for the geodesic approach) in numerical procedures. The simplest point of view is to discretize  $\Omega$  over a regular grid, and approximate integrals by sums and derivatives by finite differences. This generally leads to relatively simple expressions of the gradient, but also requires a minimization over a huge quantity of parameters.

Another approach consists in decomposing  $g$  over a basis of functions; the most popular representation is finite elements, and is studied, for example, in [54]. Finite elements generate a smooth (generally piecewise polynomial) basis of functions over which the deformation is expanded, yielding a finite number of coefficients on which the optimization is performed (see, for example . It is also possible to *define* the operator  $L$  directly from the decompositions over an orthonormal basis. Typically, an orthonormal basis  $e_1, \dots, e_n, \dots$  of  $\mathcal{L}^2$  being given,  $L$  can be defined as the operator which associates to  $u = \sum_{n=1}^{\infty} \alpha_n e_n$  the function  $Lu = \sum_{n=1}^{\infty} c(n) \alpha_n e_n$  where  $c(n)$  tends to infinity.

We can here also quote the interesting work of Pentland and Sclaroff ([51], [56]) in which the deformation dynamics (in the linear elastic formulation) is expressed in function of a finite number of coordinates by finite-element expansion. This expansions is then decomposed into *vibration modes* which are used, in particular, to characterize the shape of an object.

#### Definition of the gradient

The second remark is a fundamental fact on gradient descent. Consider first the finite dimensional case: when  $g \mapsto U(g)$  is a differentiable function on  $\mathbb{R}^p$ , the first order expansion of  $U(g+h)$  is  $U(g+h) \simeq U(g) + d_g U(h)$ , with

$$d_g U(h) = \sum_{i=1}^p \frac{\partial U}{\partial g_i} h_i$$

Gradient descent algorithms are based on the following analysis: denote by  $h.k = \sum_{i=1}^p h_i k_i$  the usual inner product on  $\mathbb{R}^p$ , and by  $|g|$  the associated norm; to determine the optimal direction of descent for  $U$  starting from  $g$ , minimize the first order expansion  $U(g) + d_g U(h)$  with the constraint that  $|h| = 1$ : it is easily shown that the optimal  $h$  has coordinates  $h_i$  proportional

to  $-\frac{\partial U}{\partial g_i}$ , leading to the well-know gradient descent algorithm

$$g_i(t+1) - g_i(t) = -\gamma \frac{\partial U}{\partial g_i}.$$

or, in vector form:

$$g(t+1) - g(t) = -\gamma \frac{dU}{dg}.$$

However, it should be noted that this expression is deeply related to the choice of the inner-product on  $\mathbb{R}^p$ . Any other choice, let's say  $\langle h, k \rangle = {}^t h A k$ , where  $A$  is a  $p \times p$  definite positive symmetric matrix, would yield another "optimal" direction, namely  $A^{-1} \frac{dU}{dg}$ . One can go even further in the generality, by letting the inner-product also depend on the current position,  $g$ , introducing therefore a mapping

$$g \mapsto A_g$$

which associates to every  $g \in \mathbb{R}^p$  a positive definite matrix  $A_g$ <sup>6</sup>, and fixing the optimal direction starting from  $g$  according to the inner-product  $\langle h, k \rangle_g = {}^t h A_g k$ . This yields the general gradient descent algorithm, which is given, in finite dimension, by

$$g(t+1) - g(t) = -\gamma \nabla_g U = -\gamma A_g^{-1} \frac{dU}{dg}.$$

the gradient being defined by

$$d_g U(h) = \langle \nabla_g U, h \rangle_g.$$

This point of view is also valid, at least formally in infinite dimension, as soon as an inner product is selected. However, this induces much deeper consequences on the practical implementation of the algorithm.

Consider now the case when  $g$  is a function defined on  $\Omega$ , since this is our case of interes here. Classical calculus of variations computes the gradient (and the Euler equations) according to the  $\mathcal{L}^2$  inner-product

$$\langle h, k \rangle_2 = \int_{\Omega} k h dx$$

therefore looking for a first order expansion of the kind

$$U(g+h) \simeq U(g) + \int_{\Omega} b_g \cdot h dx$$

---

<sup>6</sup>In other terms, one defines a general Riemannian metric structure on  $\mathbb{R}^p$

with a certain function  $b_g$  defined on  $\Omega$ , which represents the  $\mathcal{L}^2$  gradient, and yield the usual Euler equations. For example, if  $\Omega = ]0, 1[$  and

$$U(g) = \int_0^1 (|\dot{g}|^2 + V(g(x)))dx$$

for a positive potential  $V$  and  $g : [0, 1] \rightarrow [0, 1]$  subject to boundary conditions  $g(0) = 0$  and  $g(1) = 1$ , standard computations yield

$$U(g+h) \simeq U(g) + \int_0^1 \left(-\frac{d^2g}{dx^2} + V' \circ g\right) \cdot h dx$$

which allows to write the  $L^2$  gradient as

$$\nabla_g U = -\frac{d^2g}{dx^2} + V' \circ g.$$

and the gradient-descent algorithm would write

$$\frac{\partial g}{\partial t} = -\frac{d^2g}{dx^2} + V' \circ g$$

However, there are other choices for the inner-product, which may prove to be much more efficient to perform minimization. Return to the general case, and consider the product

$$\langle h, k \rangle = \int_{\Omega} k L h dx$$

for some positive, symmetric, differential operator  $L$ . Still assuming that  $U$  can be written as

$$U(g+h) \simeq U(g) + \int_{\Omega} b_g \cdot h dx,$$

one finds that the gradient of  $U$  at  $g$ , for the considered inner-product, should be defined as the solution of the partial differential equation  $Lv = b_g$ , which yields the so-called semi-implicit scheme

$$\begin{cases} \frac{\partial g}{\partial t} = -v \\ Lv = b_g \end{cases}$$

to which must be added possible boundary conditions on  $\partial\Omega$ .

Returning to the previous one-dimensional example, and considering the inner-product

$$\langle k, h \rangle = \int_0^1 k \dot{h} dx = \int_0^1 k L h dx$$

with  $Lh = -\frac{d^2h}{dx^2}$ , we obtain the algorithm

$$\begin{cases} \frac{\partial g}{\partial t} = -v \\ \frac{d^2v}{dx^2} = -\frac{d^2g}{dx^2} + V' \circ g \\ v(0) = v(1) = 0 \end{cases}$$

This algorithm (for the example and the general case), can be interpreted as a two step procedure, of smoothed gradient computation. Indeed, if  $G$  is the Green function for  $L$  (for the given boundary conditions), the solution of  $Lv = b_g$  is given by

$$v = \int_{\Omega} G(x, y)b_g(y)dy$$

In discrete time, the gradient descent corresponds to the following updating procedure ( $g(t)$  is the current state of the algorithm):

1. Compute the differential (the  $\mathcal{L}^2$ -gradient)  $b_{g(t)}$
2. Smooth  $b_{g(t)}$  by applying the kernel  $G$
3. Increment  $g(t)$  by a multiple of the smoothed differential to obtain  $g(t+1)$

Even more interesting, as an illustration of the advantages of using a “natural” gradient instead of the  $\mathcal{L}^2$  one, is the computation which has been led in [63]. Consider the functional

$$U(g) = \int_{\Omega} |I \circ g(x) - J(x)| dx$$

for two images  $I$  and  $J$  defined on  $\Omega$ ,  $g$  being a diffeomorphism on  $\Omega$ . The  $\mathcal{L}^2$ -gradient is readily computed, by computing the effects of the variation  $g \mapsto g + h$ , and is given by

$$b_g(x) = 2(I(g(x)) - J(x))D_{g(x)}I \tag{6}$$

An  $\mathcal{L}^2$  gradient descent would therefore write

$$g_{k+1} - g_k = \gamma(I(g_k(x)) - J(x))D_{g_k(x)}$$

This procedure is very unstable, and requires modifications (in particular, smoothing, see the Demon’s algorithm in the next paragraph). Another drawback is that it works as if  $g$  was searched for in a Hilbert space, instead of an infinite dimensional group. There is a high interest in incorporating

the group structure in the framework. We shall return later on how to devise invariant metrics on the group of diffeomorphisms. For the moment, we take as a definition the following inner product, between two functions  $h$  and  $k$  defined on  $\Omega$ :

$$\langle h, k \rangle_g = \int_{\Omega} L(h \circ g^{-1}) \cdot L(k \circ g^{-1}) dx$$

(the differential operator  $L$  is applied to the functions  $h \circ g^{-1}$  and  $k \circ g^{-1}$ ). We now try to express the approximation

$$U(g+h) - U(g) \simeq \int_{\Omega} 2(I(g(x)) - J(x)) D_{g(x)} I \cdot h(x) dx$$

under the form

$$U(g+h) - U(g) \simeq \langle \beta_g, h \rangle_g$$

We first make the change of variables  $y = g(x)$  in the integral, yielding

$$U(g+h) - U(g) \simeq \int_{\Omega} 2(I(y) - J(g^{-1}(y))) D_y I \cdot h(g^{-1}(y)) |\det Dg^{-1}(y)| dy$$

Introduce the operator  $L^*$  such that

$$\int_{\Omega} k \cdot Lh = \int_{\Omega} h \cdot L^*k$$

so that

$$\langle h, k \rangle_g = \int_{\Omega} (h \circ g^{-1}) \cdot L^* L(k \circ g^{-1}) dx$$

We see that our “natural gradient” (as this is sometimes called) must be such that  $\beta_g \circ g^{-1}$  is solution of the partial differential equation

$$L^* L.u(y) = 2 |\det Dg^{-1}(y)| (I(y) - J(g^{-1}(y))) D_y I$$

In other terms, gradient-descent for this inner-product must follows, at each iteration, the following steps:

1. Compute  $\alpha_k(y) = 2 |\det Dg_k^{-1}(y)| (I(y) - J(g_k^{-1}(y))) D_y I$
2. Solve the P.D.E  $L^* L.u_k = \alpha_k$
3. set  $g_{k+1} = g_k + \gamma \alpha_k \circ g_k$

As remarked before, step 2 is a smoothing step, and can be very efficiently performed if the inverse of  $L^*L$  is known (the Green function).

Step 3. is even more interesting: it is a discretization of the continuous time process

$$\frac{\partial g}{\partial t} = \alpha(t, g(t, x))$$

which means that  $g$  is approximately equal to the solution of  $\dot{y} = \alpha(t, y)$ . It is thus a flow associated to the ODE, and thus, under mild conditions, a diffeomorphism.

### Demons algorithm

Among these techniques (although not quite exactly based on gradient computation) an important one is Thirion's demons method ([58, 59, 33]) which can be interpreted as follows:

Consider, again, the energy

$$U(g) = \int_{\Omega} |J(x) - I(g(x))|^2 dx$$

with  $\mathcal{L}^2$ -gradient is given by (6):

$$b_g(x) = (I(g(x)) - J(x))D_{g(x)}I$$

Consider now the inner-product

$$\langle h, k \rangle_g = \int_{\Omega} (|J(x) - I(g(x))|^2 + |\nabla J(x)|^2)h(x)k(x)dx$$

then, the gradient of  $U$  for this inner-product is

$$\frac{b_g(x)}{|J(x) - I(g(x))|^2 + |\nabla J(x)|^2}$$

which is close to

$$\frac{J(x) - I(x)}{|J(x) - I(g(x))|^2 + |\nabla J(x)|^2} \nabla J(x)$$

The gradient descent algorithm can therefore be written (in discrete time)

$$u(t+1) = u(t) - \frac{J(x) - I(x)}{|J(x) - I(g(t, x))|^2 + |\nabla J(x)|^2} \nabla J(x) \quad (7)$$

where  $g(t, x) = x + u(t, x)$ . This is almost Thirion's demons algorithm, the only difference (which is important, since no regularization term has been incorporated to the energy) is that the right-hand term of (7) is smoothed by a Gaussian kernel before updating.

This smoothing step apart, it is interesting to try to understand the influence of the new inner-product; the value of  $\langle h, h \rangle_g$  must be interpreted as an *a priori* penalty for a variation  $g \rightarrow g + h$ ; so, variations  $h$  which are preferred are

- small when  $|J(x) - I(g(x))|^2 + |\nabla J(x)|^2$  is large, which implies that matching by  $g$  is wrong at  $x$  and  $J$  has large variations around  $x$
- large when the grey-level values which are matched are similar, but  $J$  is almost constant around  $x$

### Multiscale approaches

For these high dimensional, non convex minimization problems, good quality results most of the time requires a multi-resolution, coarse to fine strategy, both for increasing the speed of the algorithms, and to (hopefully) avoid local minima. Given this choice, there are various strategies which can be adopted. Coarsening procedures can be applied to the data (the matched images), or to the unknown (the diffeomorphism), or, of course, to both. Moreover, many variants are available in each case (pyramids, smoothing and subsampling for data, decomposition over bases of increasing resolution for the diffeomorphism, etc...). We shall not push further the details here, and refer to [9, 26, 10].

## 5 Defining distances

### 5.1 General facts

We now focus on how to define distances to compare objects like curves, shapes, 2D or 3D images, in a way which is related the action of deformations. We start with a short algebraic section in which basic facts on how inducing a distance from a group action are obtained, introducing in particular a "least-action principle"

A distance on a set  $\mathcal{I}$  is a mapping  $d : \mathcal{I}^2 \mapsto \mathbb{R}_+$  such that, for all  $I, I', I'' \in \mathcal{I}$ ,

$$\text{D1) } d(I, I') = 0 \Leftrightarrow I = I'.$$

$$\text{D2) } d(I, I') = d(I', I)$$

$$\text{D3) } d(I, I'') \leq d(I, I') + d(I', I'')$$

If D1) is not true and  $d(I, I) = 0$  for all  $I$ , we use the term *pseudo-distance*.

A group  $G$  is acting on  $\mathcal{I}$  if an operation is defined on  $G \times \mathcal{I}$  with values in  $\mathcal{I}$  such that  $\text{id}_G.I = I$  and  $g.(h.I) = (gh).I$  for all  $I \in \mathcal{I}$  and all  $g, h \in G$ . If  $G$  is a group acting on  $\mathcal{I}$ , one says that a distance  $d$  on  $\mathcal{I}$  is  $G$ -equivariant if and only if, for all  $g \in G$ , for all  $I, I' \in \mathcal{I}$ ,  $d(g.I, g.I') = d(I, I')$ .

A mapping  $d : \mathcal{I}^2 \mapsto \mathbb{R}_+$  is a  $G$ -invariant distance if and only if it satisfies D2) and D3) and D1') which is:

$$\text{D1')} \quad d(I, I') = 0 \Leftrightarrow \exists g \in G, g.I = I'.$$

The next proposition states that a  $G$ -equivariant distance induces, through a least action minimization, a  $G$ -invariant (pseudo-)distance:

**Proposition 1** *Let  $d$  be  $G$ -equivariant on  $\mathcal{I}$ . Define  $\tilde{d}$ , by*

$$\tilde{d}(I, I') = \inf\{d(g.I, g'.I') : g, g' \in G\} = \inf\{d(g.I, I') : g \in G\}.$$

*Then  $\tilde{d}$  is a  $G$ -invariant pseudo-distance.*

It will also be interesting to consider another type of construction: we still let  $G$  act on  $\mathcal{I}$ , but we now consider the product  $\mathcal{O} = G \times \mathcal{I}$ . The group  $G$  also acts on  $\mathcal{O}$ , simply letting, for  $k \in G$ ,  $o = (h, I) \in \mathcal{O}$ :

$$k.o = (kh, k.I).$$

For  $o = (h, I) \in \mathcal{O}$ , we let  $\pi(o) = h^{-1}.I$ . Assume that  $d_{\mathcal{O}}$  is a distance on  $\mathcal{O}$ , and let, for  $I, I' \in \mathcal{I}$

$$d(I, I') = \inf\{d_{\mathcal{O}}(o, o'), o, o' \in \mathcal{O}, \pi(o) = I, \pi(o') = I'\} \quad (8)$$

Then

**Proposition 2** *If  $d_{\mathcal{O}}$  is  $G$ -equivariant, then  $d$  in (8) is a pseudo-distance on  $\mathcal{I}$*

## 5.2 Infinitesimal approach

A standard way for building distances on sets like  $\mathcal{O}$  is to look for shortest paths. Assume that we are able to give a meaning of the speed  $V_{\mathbf{o}}(t) = \frac{d\mathbf{o}}{dt}$  of a path  $\mathbf{o} : t \mapsto \mathbf{o}(t)$  over  $\mathcal{O}$ , and that, for each  $o \in \mathcal{O}$ , we have a way to quantify the speed of paths passing through  $o$  with the help of a norm

$V \mapsto \|V\|_{\mathbf{o}}$  (the norm depends on  $\mathbf{o}$ ). Then, the associated path energy is given by

$$E(\mathbf{o}) = \int_0^1 \|V_{\mathbf{o}}(t)\|_{\mathbf{o}(t)}^2 dt \quad (9)$$

and the *geodesic distance* on  $\mathcal{O}$  is then defined by

$$d_{\mathcal{O}}(o, o') = \inf\{\sqrt{E(\mathbf{o})}, \mathbf{o}(0) = o, \mathbf{o}(1) = o'\}. \quad (10)$$

When  $\mathcal{O}$  is built as in the previous section, we have seen that there was some interest in building  $G$ -equivariant distances. In fact, to build a  $G$ -equivariant geodesic distance, to start with a family of norms  $(\|\cdot\|_o, o \in \mathcal{O})$  which shares this property, in the sense that, if  $\mathbf{o}$  is a path on  $\mathcal{O}$  and  $h \in G$ , then the translated path  $h.\mathbf{o}$  and  $\mathbf{o}$  share the same speeds at the same times: this writes

$$\|V_{h.\mathbf{o}}(t)\|_{h.\mathbf{o}(t)} = \|V_{\mathbf{o}}(t)\|_{\mathbf{o}(t)} \quad (11)$$

One can interpret this formula with the help of the differential of the action of  $G$  on  $\mathcal{O}$ , but the meaning and the consequences of (11) will always be easily derived in the next examples, and we will not need to introduce the usual machinery of differential geometry.

It is important to notice that this condition provides norms of velocities at translated objects  $h.o$  as soon as these norms are known at  $o$ . In the case when  $\mathcal{O} = G \times \mathcal{I}$ , it thus suffices to define  $\|\cdot\|_o$  for  $o \in \mathcal{O}$  of the kind  $o = (\text{id}_G, I)$ .

### 5.3 Invariant distances between landmarks

As a first illustration, we describe some aspects of the results of Kendall ([40]) concerning invariant distances between sets of landmarks, when these sets are considered up to affine similitudes (translation, rotation and homotheties).

We shall limit ourselves to the simplest 2D case. Fix an integer  $N > 0$  and denote by  $\mathcal{P}_N$  the set of all collections of  $N$  landmarks  $(z_1, \dots, z_N) \in (\mathbb{R}^2)^N$ . Two sets of landmarks,  $(z_1, \dots, z_N)$  and  $(z'_1, \dots, z'_N)$  are considered as equivalent if there exists an affine similitude  $g$  such that, for all  $i$ ,  $z'_i = g.z_i$ . This yields an equivalence relation between sets of landmarks, and the equivalence classes have been called  $N$ -shapes in [40].

To write simpler formulas, it will be better to identify the set  $\mathbb{R}^2$  to the set  $\mathcal{C}$ , a point  $z = (x, y)$  being associated to  $z = x + iy$ . With this notation, an affine similitude takes the form  $z \mapsto a.z + b$ , with  $a, b \in \mathcal{C}, a \neq 0$ .

For  $Z = (z_1, \dots, z_N) \in \mathcal{P}_N$ , let  $\bar{z}$  be the center of inertia

$$\bar{z} = (z_1 + \dots + z_N)/N$$

and let  $S(Z)^2 = \sum_{i=1}^N |z_i - \bar{z}|^2$ .

Let  $\Sigma_N$  be the set of  $N$ -shapes. An element of  $\Sigma_N$  will be denoted as  $[Z]$  which means  $Z \in \mathcal{P}_N$  seen up to affine similitudes. To define a distance between two classes  $[Z]$  and  $[Z']$ , we use proposition 1, letting  $G$  be the group of similitudes and  $\mathcal{I} = \mathcal{P}_N$ , so that it suffices to build a  $G$ -equivariant distance. This will be done via the infinitesimal approach.

According to the previous section, we must define a norm at each point  $Z \in \mathcal{P}_N$ , which satisfies (11) for every affine similitude  $g$ , and for every paths  $O$  in  $\mathcal{P}_N$ . First remark that, since  $\mathcal{P}_N$  is a finite dimensional vector space, the notion of velocity is trivially defined, and, since affine similitudes are linear, one has  $V_{aO+b} = aV_O$ . This implies that condition (11) is in this case equivalent to: for all  $V = (v_1, \dots, v_N) \in \mathcal{C}^N$ , for all  $Z \in \mathcal{P}_N$  and for all  $a, b \in \mathcal{C}, a \neq 0$ :

$$\|V\|_Z = \|a.V\|_{a.Z+b}$$

A consequence of this is that the norm has to be defined only for sets of landmarks  $Z$  such that  $S(Z) = 1$  and  $\bar{z} = 0$ , since, for all  $Z$

$$\|V\|_Z = \left\| \frac{1}{S(Z)} \cdot V \right\|_{\frac{Z-bZ}{S(Z)}} \quad (12)$$

So, let us make the following choice: for  $\bar{z} = 0$  and  $S(Z) = 1$ ,

$$\|V\|_Z = \sum_{i=1}^N |v_i|^2$$

Translating the norms according (12), we can compute the energy of a path  $t \mapsto W(t) \in \mathcal{P}_N$ , which is given by

$$E(W) = \int_0^1 \frac{\sum_{i=1}^N |\dot{W}_i(t)|^2}{\sum_{i=1}^N |W_i(t) - \bar{w}(t)|^2} dt$$

To compute the distance between two sets of landmarks  $Z$  and  $Z'$  this energy must be minimized over all paths  $W(\cdot)$  with  $W(0) = Z$  and  $W(1) = Z'$ . To perform this computation, introduce new variables  $v_i(t) = (W_i(t) -$

$\bar{w}(t)/\|W(t)\|$  and  $\rho(t) = S(W(t))$ . The path  $W(\cdot)$  is uniquely determined by the triple  $(\mathbf{v}(\cdot), \rho(\cdot), \bar{w}(\cdot))$ . Moreover

$$\frac{dW_i}{dt} = \frac{d\bar{w}}{dt} + \rho \frac{dv}{dt} + v \frac{d\rho}{dt}$$

and the energy writes

$$E(W) = \int_0^1 \sum_{i=1}^N \left| \frac{1}{\rho} \frac{d\bar{w}}{dt} + \frac{1}{\rho} \frac{d\rho}{dt} \cdot v_i + \frac{dv_i}{dt} \right|^2 dt$$

which can be further simplified in<sup>7</sup>

$$E(W) = N \int_0^1 \left( \frac{1}{\rho} \frac{d\bar{w}}{dt} \right)^2 + \int_0^1 \left( \frac{1}{\rho} \frac{d\rho}{dt} \right)^2 dt + \int_0^1 \sum_{i=1}^N \left| \frac{dv_i}{dt} \right|^2 dt.$$

The last term in this expression only involves  $v$  and can be explicitly minimized, with the constraints  $\sum_i v_i = 0$  and  $\sum_i |v_i|^2 = 1$ : these constraints indeed imply that  $v$  belongs to a sphere of dimension  $N - 1$ , on which the geodesic distances are given by the lengths of large circles, so that the minimum of the last term is  $\arccos(\langle \mathbf{v}(0), \mathbf{v}(1) \rangle)^2$ .

We thus obtain

$$\begin{aligned} D(Z, Z')^2 &= \inf \left( N \int_0^1 \left( \frac{d\bar{w}}{dt} \right)^2 + \int_0^1 \left( \frac{d\rho}{dt} \right)^2 dt \right) \\ &\quad + \arccos \left( \left\langle \frac{Z - \bar{z}}{S(Z)}, \frac{Z' - \bar{z}'}{S(Z')} \right\rangle \right)^2 \end{aligned}$$

the first infimum being taken over all functions  $t \mapsto (\bar{w}(t), \rho(t)) \in \mathcal{C} \times [0, +\infty[$ , such that  $\bar{w}(0) = \bar{z}$ ,  $\bar{w}(1) = \bar{z}'(1)$ ,  $\rho(0) = S(Z)$ ,  $\rho(1) = S(Z')$ .

The induced distance on  $\Sigma_N$  is now given by

$$d([Z], [Z']) = \inf \{ D(Z, aZ' + b), a, b \in \mathcal{C} \}$$

One can always cancel the terms in  $\bar{w}$  and  $\rho$  by picking  $a$  and  $b$  such that  $S(Z) = S(aZ' + b)$  and  $\bar{z} = a\bar{z}' + b$ : this fixes the values of  $b$  and of  $|a|$ . But the term in  $\arccos$  is only affected by the ratio  $a/|a|$  when  $Z'$  is replaced by  $aZ' + b$ , and this ratio can be fixed independently of  $b$  and  $|a|$ ; this implies that one can write

$$d([Z], [W]) = \inf_{\eta \in \mathcal{C}, |\eta|=1} \left[ \arccos \left( \left\langle \eta \frac{W - \bar{w}}{S(W)}, \frac{Z - \bar{z}}{S(Z)} \right\rangle \right) \right]$$

---

<sup>7</sup>One uses  $\sum_i v_i = 0$  and  $\sum_i |v_i|^2 = 1$ , and the derivatives of these formulas

which yields, after a simple computation

$$d([Z], [W]) = \arccos \left| \left\langle \frac{W - \bar{w}}{S(W)}, \frac{Z - \bar{z}}{S(Z)} \right\rangle \right| \quad (13)$$

#### 5.4 Invariant distances on groups of diffeomorphisms

We now deal with the particular case when  $\mathcal{O} = G$ , and  $G$  is a group of diffeomorphisms of a set  $\Omega$ . We follow the lines of section 5.2 to build a geodesic distance on  $G$  which is invariant by the action of  $G$  on itself. This section has its own interest, but is also a motivation for the next section in which the issue of comparing object acted on by groups of diffeomorphisms is studied.

A rigorous exposition of the subject being beyond the scope of this tutorial (see [66]), we only provide formal, or heuristic arguments. We define on  $G$  the product

$$gh = h \circ g$$

If  $t \mapsto g(t) = g(t, \cdot)$  is a time dependent diffeomorphism,  $V_g(t) = \frac{\partial g}{\partial t}$  is a vector field, ie. a function defined on  $\Omega$  with value in  $\mathbb{R}^k$ . Following the lines of section 5.2, we must define, for each  $g \in G$ , a norm  $V \mapsto \|V\|_g$  defined for any vector fields  $V$  on  $\Omega$ , such that, for any path  $t \mapsto g(t)$  on  $G$  and for any  $h \in G$ , identity (11) is true, that is

$$\|V_{hg}(t)\|_{hg(t)} = \|V_g(t)\|_{g(t)}$$

But we have, by definition of the group product,  $(hg)(t, x) = g(t, h(x))$  which implies that the velocity at  $x \in \Omega$ ,  $V_{hg}(t, x)$  is simply given by  $V_g(t, h(x))$ . The fact that (11) is true for all paths implies that, for all vector field  $V$  on  $\Omega$ , one has

$$\|V \circ h\|_{hg} = \|V\|_g$$

which is equivalent to, for all  $g$

$$\|V\|_g = \|V \circ g^{-1}\|_{\text{id}}$$

This implies that only one norm has to be defined (the one indexed by the identity), and that all other norms can be deduced from it. The energy of a path  $t \mapsto g(t, \cdot)$  is given by

$$E(g) = \int_0^1 \|V_g(t) \circ g(t)^{-1}\|_{\text{id}}^2 dt$$

Set  $v_g(t, y) = V_g(t, g(t)^{-1}(y))$ , which is the velocity in Eulerian coordinates: the energy simply writes

$$E(g) = \int_0^1 \|v_g(t)\|_{\text{id}}^2 dt$$

This is exactly the kind of energy which has been derived by different arguments in section 2.4. It is interesting enough that minimizing this energy with given boundary conditions for  $g(0)$  and  $g(1)$  provides a distance.

## 5.5 Mixing deformations and object variations: landmark matching

We now return to the general point of view of section 5.1, in which we let  $G$  be a group of diffeomorphisms of  $\Omega$  and  $\mathcal{I}$  be the set of all collections of  $N$  landmarks on  $\Omega$ . An element of  $\mathcal{I}$  is thus a  $N$ -uple  $I = (p_1, \dots, p_N) \in \Omega^N$ . We still use on  $G$  the product  $gh = h \circ g$  and define the action of  $G$  on  $\mathcal{I}$  to be

$$g.I = (g^{-1}(p_1), \dots, g^{-1}(p_N))$$

which does provide a left-action:  $(gh).I = g.(h.I)$ .

To build a distance on  $\mathcal{I}$ , we define a distance on  $\mathcal{O} = G \times \mathcal{I}$  which is  $G$ -equivariant. This will be done again along the lines of section 5.2. A path in  $\mathcal{O}$  takes the form

$$o(t) = (g(t, \cdot), p_1(t), \dots, p_N(t))$$

where  $g(t, \cdot)$  is a time dependent diffeomorphism and  $p_i(t)$  is a curve in  $\Omega$  for  $i = 1, \dots, N$ . The velocity at  $o(t)$  is

$$V_o(t) = (V_g(t), \dot{p}_1(t), \dots, \dot{p}_N(t))$$

and from (11), we must have

$$\|V_o(t)\|_{o(t)} = \|V_{g^{-1}o}(t)\|_{g^{-1}o(t)}$$

so that it is only necessary to define norms at elements  $o = (id, p_1, \dots, p_N)$ . We have, with the same notation as in the previous section,

$$V_{g^{-1}o}(t) = (v_g(t), D_{p_1(t)}g(t)\dot{p}_1(t), \dots, D_{p_N(t)}g(t)\dot{p}_N(t))$$

where  $D_p g(t)$  is the differential of  $g(t, \cdot)$  with respect to spatial coordinates, evaluated at point  $p \in \Omega$ . Making the change of variables  $q_i(t) = g(t, p_i(t))$ , this is written

$$V_{g^{-1}o}(t) = (v_g(t), \dot{q}_1(t) + v_g(t, q_1(t)), \dots, \dot{q}_N(t) - v_g(t, q_N(t)))$$

and the energy of the path  $o(t)$  takes the form

$$E(o) = \int_0^1 \|(v_g(t), \dot{q}_1(t) - v_g(t, q_1(t)), \dots, \dot{q}_N(t) + v_g(t, q_N(t)))\|^2 dt$$

A simple example is to take

$$E(o) = \int_0^1 \int_{\Omega} |Lv_g(t)|^2 dt dx + \sum_{i=1}^N \int_0^1 |\dot{q}_i(t) - v_g(t, q_i(t))|^2 dt$$

To compute the distance between two elements of  $\mathcal{O}$ , it suffices to minimize the energy of the paths which link them. However, we are interested by the distance between two sets of landmarks  $I$  and  $I'$ , which is given, according to (8)

$$d(I, I') = \inf\{d_{\mathcal{O}}(o, o'), o, o' \in \mathcal{O}, \pi(o) = I, \pi(o') = I'\}$$

where  $\pi(g, p_1, \dots, p_N) = g^{-1}(p_1, \dots, p_N) = (g(p_1), \dots, g(p_N))$ . It is not difficult to check ([48]) that in fact,  $d(I, I')$  is the infimum of

$$\int_0^1 \int_{\Omega} |Lv(t)|^2 dt dx + \sum_{i=1}^N \int_0^1 |\dot{q}_i(t) - v(t, q_i(t))|^2 dt$$

over all time dependent vector fields  $v$  on  $\Omega$ , and over all curves  $q_1(\cdot), \dots, q_N(\cdot)$  such that  $I = (q_1(0), \dots, q_N(0))$  and  $I' = (q_1(1), \dots, q_N(1))$ . We thus obtain a new landmark-based matching formula, which only involves the velocity, and which, in the same time, provides a distance between sets of landmarks. Note that, for fixed trajectories  $q_i(\cdot), i = 1, \dots, N$ , the optimal  $v$  can be explicitly computed at each time  $t$ , in function of the Green kernel of  $L$ , as developed in section 3.3.

## 5.6 Mixing deformations and object variations: image matching

Exactly the same approach can be applied when the set  $\mathcal{I}$  is a set of mappings (images)  $I : \Omega \rightarrow \mathbb{R}^d$ . Now, the action is defined by  $g.I = I \circ g$ , and if  $o(t) = (g(t, \cdot), I(t, \cdot))$  is a path on  $\mathcal{O} = G \times \mathcal{I}$ , we find that the energy of  $o$  must take the form

$$E(o) = \int_0^1 \left\| \left( v_g, \frac{\partial J}{\partial t} + DJ.v \right) \right\|_{\text{id}, J}^2 dt$$

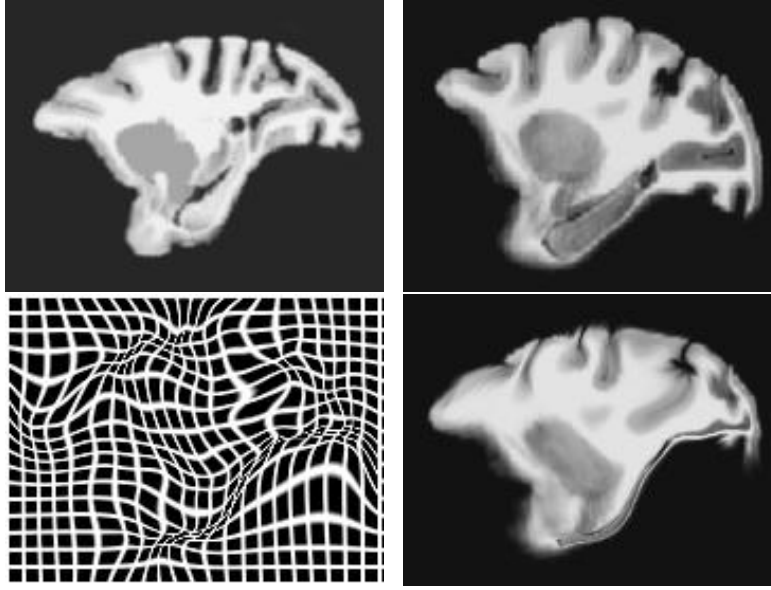


Figure 6: Matching of two macaque brain slices with (14); up: left and right: original images; down: deformed grid and application of the deformation to the upper right image

with  $J(t, y) = I(t, g(t)^{-1}(y))$ . A particular example can be

$$E(o) = \int_0^1 \int_{\Omega} |Lv|^2 dxdt + \lambda \int_0^1 \int_{\Omega} \left| \frac{\partial J}{\partial t} + DJ.v \right|^2 dxdt$$

Moreover, it can be shown that the distance  $d(I, I')$  can be computed by minimizing

$$\int_0^1 \int_{\Omega} |Lv|^2 dxdt + \lambda \int_0^1 \int_{\Omega} \left| \frac{\partial J}{\partial t} + DJ.v \right|^2 dxdt \quad (14)$$

in  $v$  and  $J$  with boundary conditions  $J(0, \cdot) = I(\cdot)$  and  $J(1, \cdot) = I'(\cdot)$ . An example of matching performed with this method is given in figure 6.

## References

- [1] F. ABRAMOVICH, T. SAPATINAS, AND W. SILVERMAN, B, *Wavelet thresholding via a bayesian approach*, J. Royal Stat. Soc. B, (1998).

- [2] A. ADAMS, R., *Sobolev Spaces*, Academic Press, 1975.
- [3] Y. AMIT, U. GRENANDER, AND M. PICCIONI, *Structural image restoration through deformable templates*, JASA, 86 (1989), pp. 376–387.
- [4] Y. AMIT AND P. PICCIONI, *A non-homogeneous markov process for the estimation of gaussian random fields with non-linear observations*, Ann. of Proba., 19 (1991), pp. 1664–1678.
- [5] N. ARAD, N. DYN, D. REISFELD, AND Y. YESHURUN, *Image warping by radial basis functions: application to facial expressions*, CVGIP: Graphical Models and Image Processing, 56 (1994), pp. 161–172.
- [6] N. ARAD AND D. REISFELD, *Image warping using few anchor points and radial functions*, Computer Graphics forum, 14 (1995), pp. 35–46.
- [7] V. ARNOL'D, *Méthodes mathématiques de la mécanique classique*, MIR, Moscow, 1976.
- [8] R. BAJCSY AND C. BROIT, *Matching of deformed images*, in The 6th international conference in pattern recognition, 1982, pp. 351–353.
- [9] R. BAJCSY AND S. KOVACIC, *Multiresolution elastic matching*, Comp. Vision, Graphics, and image proc., 46 (1989), pp. 1–21.
- [10] —, *Multiscale/multiresolution representations*, in Brain Warping, W. Toga, A, ed., 1999, pp. 45–66.
- [11] M. BAKIRCIOGLU, U. GRENANDER, N. KHANEJA, AND M. MILLER, *Curve matching on brain surfaces using induced frénet distances matrices*, Special issue of Human brain mapping (to appear), (2000).
- [12] R. BASRI, L. COSTA, D. GEIGER, AND D. JACOBS, *Determining the similarity of deformable shapes*, in IEEE Workshop on Physics based Modeling in Computer Vision, 1995, pp. 135–143.
- [13] F. BEG, M, I. MILLER, M, T. RATNANATHER, AND L. YOUNES, *On the variational equations of computational anatomy*, tech. rep., Center for Imaging sciences, John Hopkins university, 2000.
- [14] P. BELHUMEUR, J. HESPANHA, AND D. KRIEGMAN, *Eigenfaces vs. fisherfaces: reconstruction using class specific linear projection*, IEEE Trans. PAMI, (1997).

- [15] L. BOOKSTEIN, F, *Morphometric tools for landmark data; geometry and biology*, Cambridge University press, 1991.
- [16] M. BRO-NIELSEN AND C. GRAMKOW, *Fast fluid registration of medical images*, in Proceedings of VBC'96, Lecture notes in Comp. Science 1131, Springer, 1996, pp. 267–276.
- [17] G. BROWN, L, *A survey of image registration techniques*, ACM computing surveys, 24 (1992), pp. 325–376.
- [18] E. CHRISTENSEN, G, D. RABBITT, R, AND I. MILLER, M, *Deformable templates using large deformation kinematics*, IEEE trans. Image Proc., (1996).
- [19] I. COHEN, N. AYACHE, AND P. SULGER, *Tracking points on deformable objects using curvature information*, in Computer Vision - ECCV'92, G. Sandini, ed., 1992, pp. 458–466.
- [20] T. COOTES, C. TAYLOR, D. COOPER, AND J. GRAHAM, *Active shape models: their training and application*, Comp. Vis. and Image Understanding, 61 (1995), pp. 38–59.
- [21] R. DOURSAT, W. KONEN, M. LADES, C. VON DER MALSBERG, J. VORBRÜGGEN, L. WISKOTT, AND R. WÜRTZ, *Neural mechanisms of elastic pattern matching*, in Proceedings of a BMFT workshop, October 1992.
- [22] R. DOWNIE, T, L. SHEPSTONE, AND W. SILVERMAN, B, *a wavelet approach to deformable templates*, in Image fusion and shape variability techniques, V. Mardia, K, A. Gill, C, and L. Dryden, I, eds., Leeds University Press, 1996, pp. 163–169.
- [23] P. DUPUIS, U. GRENANDER, AND M. MILLER, *Variational problems on flows of diffeomorphisms for image matching*, Quaterly of Applied Math., (1998).
- [24] C. GEE, J, *On matching brain volumes*, tech. rep., Dept. of neurology, Univ. of Pennsylvania, 1998.
- [25] C. GEE, J, A. FABELLA, B, I. FERNANDES, B, I. TURETSKY, B, C. GUR, R, AND E. GUR, R, *New experimental results in atlas-based brain morphometry*, in SPIE Medical imaging 1999, 1999.

- [26] C. GEE, J. R. HAYNOR, D. L. LE BRIQUER, AND Z. BAJCSY, R, *Advances in elastic matching theory and its implementation*, in CVRMed-MRCAS'97, P. Cinquin, R. Kikinis, and D. Lavalée, eds., Springer Verlag, 1997.
- [27] C. GEE, J AND D. PERALTA, P, *Contuum models for bayesian image matching*, in Maximum entropy and bayesian methods, H. Hanson, K and N. Silver, R, eds., Kluwer academics, 1995.
- [28] D. GEIGER, A. GUPTA, A. COSTA, L, AND J. VLONTZOS, *Dynamic programming for detecting, tracking and matching deformable contours*, IEEE PAMI, 17 (1995), pp. 295–302.
- [29] A. GLASBEY, C AND V. MARDIA, K, *A review of image-warping methods*, J. of Applied Stat., 25 (1998), pp. 155–171.
- [30] W. GORMAN, J, R. MITCHELL, AND P. KUEL, F, *Partial shape recognition using dynamic programming*, IEEE PAMI, 10 (1988).
- [31] U. GRENANDER, *General Pattern Theory*, Oxford Science Publications, 1993.
- [32] U. GRENANDER AND D. M. KEENAN, *On the shape of plane images*, Siam J. Appl. Math., 53 (1991), pp. 1072–1094.
- [33] A. GUIMON, A. ROCHE, N. AYACHE, AND J. MEUNIER, *Three-dimensional brain warping using the demons algorithm and adptive intensity corrections*, tech. rep., INIRIA Sophia-Antipolis, 1999.
- [34] L. HALLINAN, P, G. GORDON, G, L. YUILLE, A, P. GIBLIN, AND D. MUMFORD, *Two and three dimensional patterns of the face*, A K Peters, Ltd, 1999.
- [35] P. HALLINAN, *A low dimensional model for face recognition under arbitrary lighting conditions*, in Preceedings CVPR'94, 1994, pp. 995–999.
- [36] S. HELGASON, *Differential Geometry, Lie groups and Symmetric spaces*, Academic Press, 1978.
- [37] K. P. HORN, B AND G. SCHUNK, B, *Determining optical flow*, Artificial intelligence, 17 (1981), pp. 185–203.
- [38] J. JONES, M AND T. POGGIO, *Multidimensional morphable models: a framework for representing and matching object classes*, Int. J. Comp. Vision, 29 (1998), pp. 107–131.

- [39] S. JOSHI, *Large deformation diffeomorphisms and Gaussian random fields for statistical characterization of brain sub-manifolds*, PhD thesis, Sever institute of technology, Washington University, 1997.
- [40] D. G. KENDALL, *Shape manifolds, procrustean metrics and complex projective spaces*, Bull. London Math. Soc., 16 (1984), pp. 81–121.
- [41] T. KENT, J AND V. MARDIA, K, *The link between kriging and thin-plate splines*, in Probability, statistics and optimisation, P. Kelly, F, ed., John Wiley & sons, 1994, pp. 325–339.
- [42] H. LE, *Mean size-and-shapes and mean shapes: a geometric point of view*, Adv. Appl. Prob., 27 (1995), pp. 44–55.
- [43] A. MARQUES, J AND J. ABRANTES, A, *Shape alignment-optimal initial point and pose estimation*, Pattern Recognition letters, 18 (1997), pp. 49–53.
- [44] J. MARSDEN AND T. HUGUES, *The mathematical foundations of elasticity*, Prentice-Hall, 1983.
- [45] J. MEINGUET, *Multivariate interpolation at arbitrary points made simple*, J. Appliet Math. and Physics, 30 (1979), pp. 292–304.
- [46] Y. MEYER, *Wavelets and operators*, Cambridge University Press, 1992.
- [47] I. MILLER, M, C. JOSHI, S, AND E. CHRISTENSEN, G, *Large deformation fluid diffeomorphisms for landmark and image matching*, in Brain Warping, A. Toga, ed., Academic Press, 1999, pp. 115–131.
- [48] I. MILLER, M AND L. YOUNES, *Group action, diffeomorphism and matching: a general framework*, in Proceeding of SCTV 99, 1999. <http://www.cis.ohio-state.edu/szhu/SCTV99.html>.
- [49] P. OLVER, *Equivalence, Invariants and Symmetry*, Cambridge University Press, 1995.
- [50] W. PAGLIERONI, D AND K. JAIN, A, *Fast classification of discrete shape contours*, Pattern reconition, 20 (1987), pp. 583–598.
- [51] A. PENTLAND AND S. SCLAROFF, *Closed-form solutions for physically-based shape modeling and recognition*, IEEE TPAMI, 13 (1991), pp. 715–729.

- [52] A. PICAZ AND I. DINSTEIN, *Matching of partially occluded planar curves*, Pattern Recognition, 28 (1995), pp. 199–209.
- [53] M. PICCIONI, S. SCARLATTI, AND A. TROUVÉ, *A variational problem arising from speech recognition*, SIAM J. Applied Math., 58 (1998), pp. 753–771.
- [54] D. RABBITT, R. A. WEISS, J. E. CHRISTENSEN, G. AND I. MILLER, M., *Mapping of hyperelastic deformable templates using the finite element method*, in Proceeding of San Diego’s SPIE conference, 1995.
- [55] H. SAKOE AND S. CHIBA, *Dynamic programming algorithm optimization for spoken word recognition*, IEEE Trans. Accoustic, Speech and Signal Proc., 26 (1978), pp. 43–49.
- [56] S. SCLAROFF, *Modal matching: a method for describing, comparing and manipulating signals*, PhD thesis, MIT, 1995.
- [57] R. SZELISKI AND J. COUGHLAN, *Spline-based image registration*, tech. rep., Cambridge Research Laboratory, 1994.
- [58] J.-P. THIRION, *Image matching as a diffusion process: an analogy with maxwell’s demons*, Medical Image Analysis, 2 (1998), pp. 243–260.
- [59] ———, *Diffusing models and applications*, in Brain Warping, W. Toga, A, ed., 1999, pp. 144–155.
- [60] J.-P. THIRION AND G. CALMON, *Deformation analysis to detect and quantify active lesions in 3d medical image sequences*, IEEE Trans. Image Analysis, 18 (1999), pp. 429–442.
- [61] M. THOMSON, P AND W. TOGA, A, *Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations*, Medical Image Analysis, 1 (1996/7), pp. 271–294.
- [62] W. TOGA, A, ed., *Brain warping*, Academic Press, 1999.
- [63] A. TROUVÉ, *Diffeomorphism groups and pattern matching in image analysis*, Int. J. of Comp. Vis., 28 (1998), pp. 213–221.
- [64] A. TROUVÉ AND L. YOUNES, *Diffeomorphic matching in 1d: designing and minimizing matching functionals*, in Proceedings of ECCV 2000, D. Vernon, ed., 2000.

- [65] ———, *On a class of optimal matching problems in 1 dimension*, Siam J. Control Opt. (to appear), (2000).
- [66] A. TROUVÉ, *Infinite dimensional group action and pattern recognition*, Quaterly of Applied Math. (to appear), (2000).
- [67] M. TURK AND A. PENTLAN, *Eigenfaces for recognition*, J. of Cognitive Neuroscience, 3 (1991).
- [68] T. VETTER AND T. POGGIO, *Linear object classes and image synthesis from a single example image*, IEEE trans. PAMI, 19 (1997), pp. 733–742.
- [69] L. WISKOTT AND C. VON DER MALSBERG, *A neural system for the recognition of partially occluded objects in cluttered scenes*, Int. J. Pattern recogn. and Art. Intel., 7 (1993), pp. 935–948.
- [70] K. YOSIDA, *Functional analysis*, Springer, 1970.
- [71] L. YOUNES, *Computable elastic distances between shapes*, SIAM J. Appl. Math, 58 (1998), pp. 565–586.
- [72] A. L. YUILLE, *Generalized deformable models, statistical physics, and matching problems*, Neural computations, (1990), pp. 1–24.