11

Mean Pattern Estimation of Images Using Large Diffeomorphic Deformations

Jérémie Bigot and Sébastien Gadat
Institut de Mathématiques de Toulouse, Université Paul Sabatier (Toulouse)
France

1. Introduction

This work deals with the problem of estimating the statistical object called " mean" concerning a situation where one observe noisy images which are also corrupted through a deformation process. The difficulty of this statistical problem arises from the nature of the space in which the object to estimate are living. A popular approach in image processing is Grenander's pattern theory described in Grenander (1993) and Grenander & Miller (2007) where natural images are viewed as points in an infinite dimensional manifold and the variations of the images are modelled by the action of deformation groups on the manifold.

In the last decade, the construction of deformation groups to model the geometric variability of images, and the study of the properties of such deformation groups has been an active field of research: one may refer for instance to the works of Beg et al. (2005), Joshi et al (2004), Miller & Younes (2001) or Trouvé & Younes (2005b). But to the best of our knowledge, few results on statistical estimation using such deformations groups are available. In this setting, there has been recently a growing interest on the problem of defining an empirical mean of a set of random images using deformation groups and non-Euclidean distances. A first attempt in this direction is the statistical framework based on penalized maximum likelihood proposed in Glasbey & Mardia (2001). Computation of empirical Fréchet mean of biomedical images is discussed in Joshi et al (2004). More recently, Allassonière et al. (2007), Allassonière et al. (2009) and Ma et al. (2008) have proposed a statistical approach using Bayesian modelling and random deformation models to approximate the mean and main geometric modes of variations of 2D or 3D images in the framework of small deformations.

Starting from Bigot et al. (2009), we focus in this text on the problem of estimating a mean pattern in shape invariant models for images using random diffeomorphic deformations in two-dimensions. The main goal of this paper is to show that the framework proposed in Bigot et al. (2009) leads to the definition of an empirical Fréchet mean associated to a new dissimilarity measure between images. We study both theoretical and numerical aspects of such an empirical mean pattern. In this extended abstract, we present the main ideas of such an approach. We also give a numerical example of mean pattern estimation from a set of faces to illustrate the methodology.

Our work will be organised as follows. We first define the mathematical objects used and then model our situation of noisy and warped images with random large deformations. We aim to solve theoretically the problem of estimating the mean pattern of a set of images and state some convergence results among some general assumptions. At last, we propose an algorithm to approach the theoretical estimator and illustrate our method on real datasets.

2. Statistical deformable framework and statement of the mean pattern estimation problem

We first describe our random model of deformations on 2-dimensional images but note that this model can be extended to higher dimensions with very minor modifications. Thus, the following paragraphs are organised as follow, we first recall in paragraph 2.1.1 the classical model of large deformation introduced in the works of Younes, (2004). Next, we describe precisely the parametric decomposition of our diffeomorphism (paragraph 2.1.3) and then explain how one can use our method to generate random large deformations with localised effects.
At last, this section ends with the paragraph 2.2 and the description of the statistical model we consider for randomly warped image corrupted by additive noise.

2.1 Mathematical model of large deformation

2.1.1 Definition

For sake of simplicity, images are considered to be functions from a compact set denoted Ω which will be set equal to Ω in the sequel. Moreover, we deal in this work with grey levelled thus the generic notation for images I will be $I : \Omega \mapsto \mathbb{R}$. The first task is to define a suitable notion of deformation of the domain Ω. For this, we will adopt the large deformation model governed by diffeomorphic flows of differential equation introduced in Trouvé & Younes (2005a). These deformations denoted Φ will later be combined with some pattern of reference $I^\star$ to produce our noisy and wrapped images $I^\star \circ \Phi + \epsilon$.
Let us first describe precisely our model to generate diffeomorphisms Φ of Ω.
We first consider any smooth vector field v from Ω to $\mathbb{R}^2$ with a vanishing assumption on the boundary of Ω:

$$\forall x \in \partial\Omega \qquad v(x) = 0. \tag{1}$$

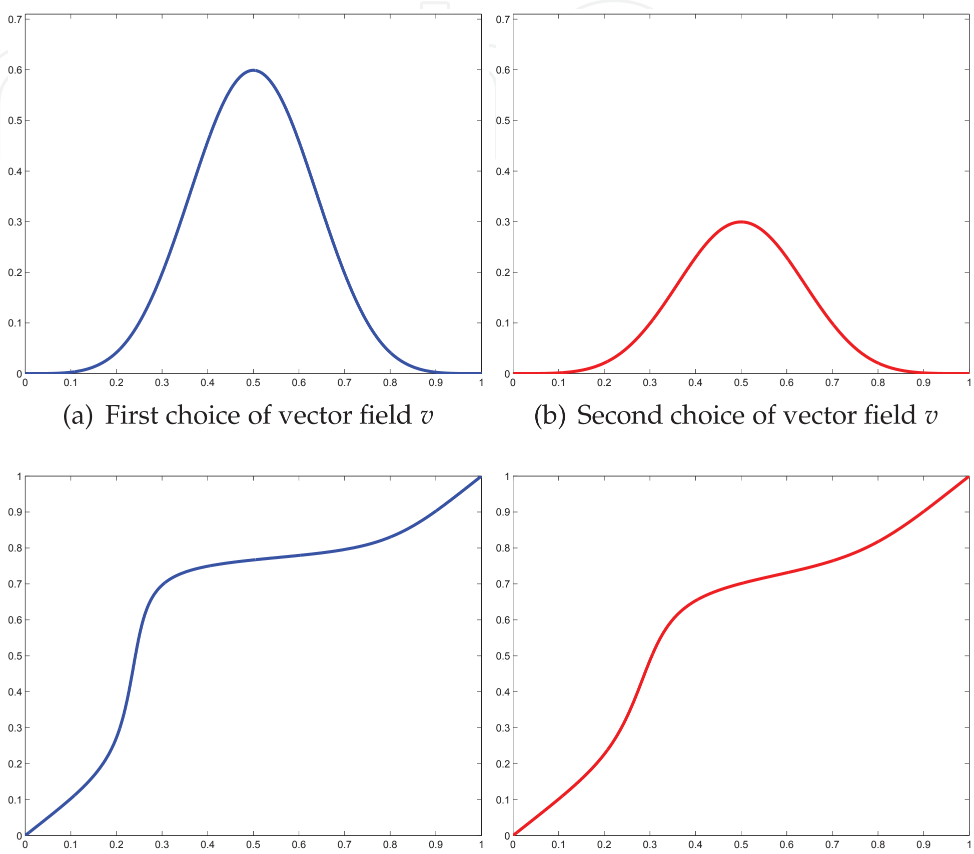
Now, we consider the set of applications $(\Phi_v^t)_{t\in[0;1]}$ from Ω to Ω, solution of the ordinary differential equation

$$\begin{cases} \forall x \in \Omega \quad & \Phi_v^0(x) = x, \\ \forall x \in \Omega \quad \forall t \in [0;1] \quad & \frac{d\Phi_v^t(x)}{dt} = v(\Phi_v^t(x)). \end{cases} \tag{2}$$

A remarkable mathematical point for the ordinary differential equation (2) is that it builds a set of diffeomorphisms on Ω, Φ_v^t for any $t \in [0;1]$.
As we want to have a deformation which remains in Ω, we have imposed that $\Phi^1_{v|\partial\Omega} = Id$, meaning that our diffeomorphism is the identity at the boundaries of Ω. Note that in the above definition, v is an homogeneous vector field (it does not depend on time t) which means that the differential equation is time-homogeneous and v is also a smooth ($C^\infty(\Omega)$) function.
The solution at time $t = 1$ denoted by Φ_v^1 of the above ordinary differential equation is a diffeomorphic transformation of Ω generated by the vector field v, which will be used to model image deformations. One can easily check that the vanishing conditions (1) on the vector field v imply that $\Phi_v^1(\Omega) = \Omega$ and that Φ_v^t is a diffeomorphism for all time $t \in [0,1]$. Thus Φ_v^1 is a convenient object to generate diffeomorphisms. Indeed, to compute the inverse diffeomorphisms of Φ_v^t, it is enough to revert the time in equation (2). One may refer to Younes, (2004) for further details on this construction.

2.1.2 One dimensional example

To illustrate the simple construction based on the choice of the vector field v and the obtained diffeomorphism Φ_v^1, we consider a first simple example in one-dimension (i.e. for $v : [0,1] \to \mathbb{R}$ which generates a diffeomorphism of the interval $[0,1]$). In Figure 1, we display two vector fields that have the same support on $[0,1]$ but different amplitudes, and we plot the corresponding deformation Φ_v^1.



(a) First choice of vector field v (b) Second choice of vector field v

(c) Diffeomorphism Φ_v^1 obtained with blue vector field v (d) Diffeomorphism Φ_v^1 obtained with red vector field v

Fig. 1. Some numerical examples of two choices for the vector field v and the obtained diffeomophisms through ordinary differential equation (2).

Note that the deformations require the prior choice of a vector field v. One can see that the amount of deformations, measured as the local distance between Φ_v^1 and the identity, depends on the amplitude of the vector field. In the intervals where v is zero, then the deformation is locally equals to the identity as pointed in Figure 1. Hence, this simple remark asserts that local deformations will be generated by choosing compactly supported vector fields which will be decomposed in localized basis functions.

2.1.3 Building the vector field v in 2D

Parametric decomposition

Consider an integer K and some linearly independent functions $e_k : \mathbb{R}^2 \to \mathbb{R}^2$ whose choices will be discussed later on. We have chosen to use vector fields v that can be decomposed on the family of functions $e_k = (e_k^1, e_k^2)$. The deformations are generated as follows. Let

(a_k^1, a_k^2), $k = 1, \ldots, K$ be coefficients in $[-A, A]$ for a given real $A > 0$. Then, we define a vector field v_a as

$$\forall x \in \Omega \qquad v_a(x) = \begin{pmatrix} \sum_{k=1}^K a_k^1 e_k^1(x) \\ \sum_{k=1}^K a_k^2 e_k^2(x) \end{pmatrix}. \tag{3}$$

Finally, one has just to run the previously defined O.D.E (2) to produce a deformation, Φ_{v_a} with a parametric representation.

Choice of e_k and numerical example

The amount of variability for diffeomorphisms Φ_{v_a} is thus related to the choice of K and basis functions e_k used to decompose the vector field v. In order to get a smooth bijection of Ω, the e_k should be at least differentiable. Such functions are built as follows. First, we choose a set of one-dimensional B-splines functions (of degree at least 2) whose supports are included in $[0; 1]$. To form two-dimensional B-splines, the common way is to use tensor products for each dimension. Recall that to define B-splines, one has to fix a set of control points and to define their degree. Further details are provided in de Boor (1978) and we will fix these parameters in the section dealing with experiments. B-splines functions are compactly supported with a local effect on the knots positions. This local influence is very useful for some problems in image warping where the deformation must be the identity on large parts of the images together with a very local and sharp effect at some other locations. The choice of the knots and the B-spline functions allows one to control the support of the vector field and therefore to define a priori the areas of the images that should be transformed.
In Figure 2, we display an example of a basis $e_k^1 = e_k^2, k = 1, \ldots, K$ for vector fields generated by the tensor product of 2 one-dimensional B-splines (hence $K = 4$). An example of deformation of the classical Lena image is shown in Figure 2 with two different sets coefficients a_k sampled from a uniform distribution on $[-A, A]$ (corresponding to different values for the amplitude A, a small and a large one). The amount of deformation depends on the amplitude of A, while the choice of the B-spline functions allows one to localize the deformation.

2.1.4 Random generation of large diffeomorphisms

Given any e_k and following our last construction, one can remark that it is enough to consider a random distribution on coefficients a to generate a large class of random diffeomorphisms. In our simulations, we take for P_A the uniform distribution on $[-A, A]$ i.e. $a_k^i \sim \mathcal{U}_{[-A,A]}$, $i = 1, 2$. However, it should mentioned that in the sequel, P_A can be any distribution on $\mathbb{R}$ provided it has a compact support. The compact support assumption for P_A is mainly used to simplify the theoretical proofs for the consistency of our estimator.

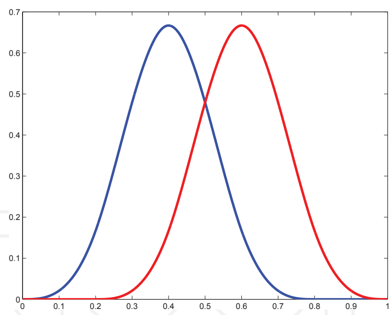
2.2 Random image warping model with additive noise

We fix discretization of Ω as a square grid of $N = N_1 \times N_2$ pixels. Given any image of reference $I^\star$ and a vector field v defined on Ω, we generate a diffeomorphism through or ordinary differential equation and we can define the general warping model by:

Definition 1 (Noisy random deformation of image)**.** *Fix an integer K and a real $A > 0$, we define a noisy random deformation of the mean template $I^\star$ as*

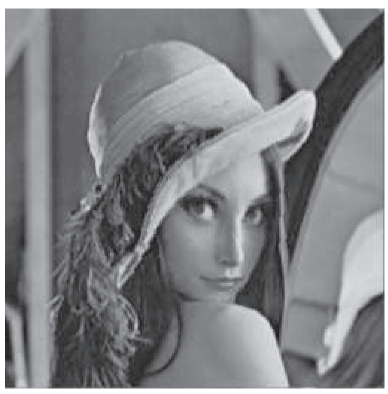
$$I_{\varepsilon,a}(p) = I^\star \circ \Phi_{v_a}^1(p) + \varepsilon(p),\ p \in \Omega,$$

where a is sampled from a distribution $P_A^{\otimes 2K}$ and ε is an additive noise independent from the coefficients a. The new image $I_{\varepsilon,a}$ is generated by deforming the template $I^\star$ (using the composition rule $\circ$) and by adding a white noise at each pixel of the image.

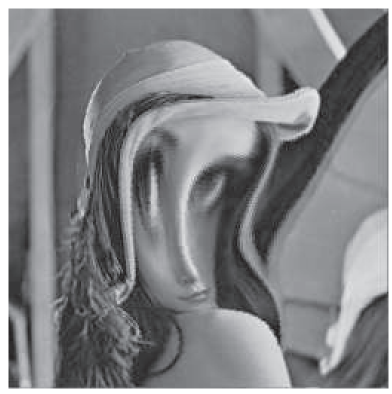


(a) Choice of 1D B-splines

(b) Original image of Lena



(c) Small deformation of Lena with a random uniform sampling of coefficients a (small value of A)



(d) Small deformation of Lena with a random uniform sampling of coefficients a (large value of A)

Fig. 2. Some numerical examples of two choices for the vector field v and the obtained diffeomophisms through ordinary differential equation (2).

Remark 1. *In our theoretical approach, we consider the pixels p as a discretization of the set Ω since our applications will be set up in this framework. However, it may be possible to handle this model in a continuous setting using the continuous white noise model. This model involves the use of an measure integration over Ω instead of sums over the pixels p of the image and we refer to Candes & Donoho (2000) for further details. Finally, remark that the image $I^\star$ is considered as a function of the whole square Ω, giving sense to $I^\star \circ \Phi_u^1(p), \forall p \in \Omega$.*

For notation convenience, we will denote Φ_a the diffeomorphism obtained at time $t = 1$ with the ordinary differential equation (2) based on the vector field v_a, $\Phi_a = \Phi_{v_a}^1$.

2.3 Statement of the problem and mathematical assumption

Using this definition of randomly warped image with additive noise, we consider now a set of n noisy images that are random deformations of the same unknown template $I^\star$ as follows:

$$\forall p \in \Omega \qquad I_{a^i,\varepsilon^i}(p) = I^\star \circ \Phi_{a^i}^1(p) + \varepsilon^i(p),\ i = 1, \ldots, n. \tag{4}$$

where ε^i are i.i.d unknown observation noise and a^i are i.i.d unknown coefficients sampled as $P_A^{\otimes K \times n}$. **Our goal is as well to estimate the mean template image $I^\star$ as to infer some several**

applications of this estimation to image processing.

One may directly realize that this problem is not so easy since the space where the denoised versions of I_{a^i,ε^i} are leaving is not a classical euclidean space such as $\mathbb{R}^d$ since the diffeomorphisms that generate the image changes from one sample i to another one i'. For instance, consider 10 handwritten realisations of the digit "two", a simple arithmetic mean does not yield satisfactory results as pointed in Figure 3.

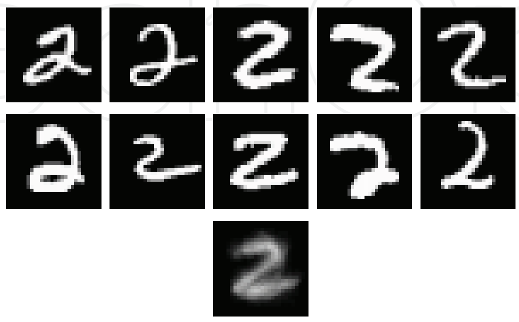


Fig. 3. Naive mean (two first rows) of a set of 10 images (Mnist database, 28×28 pixels images, see LeCun et al. (1998) for more details on this data set) and naive arithmetic mean of these images.

For our theoretical study, we will need some mathematical assumptions:

A1 There exists a constant C such that

$$|\varepsilon| < C.$$

A2

$$I^\star \text{is L-Lipschitz.}$$

Assumption **A1** means that the level of noise is bounded which seems reasonable since we generally observe gray-level images which take values on a finite discrete set.

Assumption **A2** is more questionable since a direct consequence is that $I^\star$ is continuous, which seems impossible for natural models of images with structural discontinuities (think of the space of bounded variation (BV) functions for instance). However, one can view $I^\star$ as a map from all points in Ω rather than just a function defined on the pixels. On Ω, it is more likely to suppose that $I^\star$ is the result of the convolution of $\mathcal{C}^\infty$-filters with captors measurements, which yields a smooth differentiable map on Ω. One may see for instance the work of Faugeras & Hermosillo (2002) for further comments on this assumption.

3. Statistical estimation of a mean pattern

3.1 The statistical problem

Consider a set of n noisy images $I_1, \ldots I_n$ that are independent realizations from the model (4) with the same original pattern $I^\star$. We first aim at constructing an estimate of this pattern of reference $I^\star$. We are looking at an algorithm that becomes sharper and sharper around the true pattern $I^\star$ when the number of observations n goes to $+\infty$. Without any convex structure on the images, averaging directly the observations is likely to blur the n images without yielding a sharp "mean shape". Indeed, computing the arithmetic mean of a set of images to estimate the mean pattern does not make sense as the space of deformed images $I^* \circ$

Φ_v^1 and the space of diffeomorphisms are not vectorial spaces. This is illustrated in the former paragraph in Figure 3. To have a consistent estimation of $I^\star$, one needs to solve an inverse problem as stated in the works of Biscay & Mora (2001) and Huilling (1998) derived from the random deformable model (4) since it is needed to remove the random warping effect (and consequently invert each diffeomorphism $\Phi_{a^i}^1$ which are unknown) before removing the noise ε^i on each image. Thus, recovering $I^\star$ is not an obvious task which requires some sophisticated statistical tools. A short description is provided in the next paragraph.

3.2 Presentation of M-estimation techniques

We will consider an optimisation problem (minimize an energy in our case) whose solution is the pattern $I^\star$ we are looking for and we will use some classical M-estimation techniques. Intuitively the ideas underlying M-estimation in statistics (see van der Waart (1998)) can be understood by considering the following simple example provided for a better understanding.

Example 1. *Let $X_1, \dots X_n \sim_{i.i.d.} P$ and $\alpha^* = \mathbb{E}_P[X_1]$. The simplest way to estimate α^* is $\hat{\alpha_n} = \frac{1}{n}(X_1 + \cdots + X_n)$. If for some real α, we define F_n and F the functions as*

$$F_n(\alpha) = \sum_{i=1}^{n}(X_i - \alpha)^2 \quad \text{and} \quad F(\alpha) = \mathbb{E}_P[(X-\alpha)^2],$$

then one can easily check that $\hat{\alpha}_n$ is the minimum of F_n and that $\alpha^\star$ is the minimum of F. Of course, F is unknown since it depends on the unknown law P, but a stochastic approximation of F is provided by F_n as soon as one observe $X_1, \dots X_n$. Moreover, one can remark that $F_n(\alpha) \to F(\alpha)$ almost surely (a.s.) as $n \to \infty$.
Thus, a simple way to obtain an estimate of $\alpha^\star = \arg\min F$ is to remark that the minimum of F_n should concentrates itself around the minimum of F as n is going to $+\infty$. Mathematically, this is equivalent to establish some results like:

$$F_n \longmapsto_{n\to\infty} F \Longrightarrow \arg\min F_n \longmapsto_{n\to\infty} \arg\min F.$$

In our framework, estimating the pattern I^* involves finding a best image that minimizes an energy for the best transformation which aligns the observations onto the candidate. So we will therefore define an estimator of $I^\star$ as a minimum of an empirical contrast function F_n (based on the observations $I_1, \dots I_n$) which converges, under mild assumptions, toward a minimum of some contrast F.

3.3 A new contrast function for estimating a mean pattern

We start this paragraph with some notations.

Definition 2 (Contrast function)**.** *Denote by $\mathcal{Z} = \{Z : \Omega \to \mathbb{R}\}$ a set of images uniformly bounded (e.g. by the maximum gray-level). Then, define $\mathcal{V}_A$ as the set of vector fields given by (3). An element v_a in $\mathcal{V}$ can thus be written as*

$$v_a = \left(\sum_{k=1}^{K} a_k^1 e_k^1, \sum_{k=1}^{K} a_k^2 e_k^2 \right), \text{ for some } a_k^i \in [-A, A].$$

If we denote N the number of pixels, we define the function f as

$$f(a, \varepsilon, Z) = \min_{v \in \mathcal{V}_A} \sum_{p=1}^{N} \left(I_{a,\varepsilon}(p) - Z \circ \Phi_v^1(p) \right)^2, \tag{5}$$

where I is a given image of $\mathcal{Z}$, the vector field $v_a \in \mathcal{V}_A$.
In the expression of f, one can remark that a varies in a compact set (finite number of bounded coefficients) and consequently the definition of f makes sense. Intuitively, f must be understood as the optimal (minimum) cost to align the image Z onto the noisy randomly warped image $I_{a,\varepsilon}$ using a diffeomorphic transformation.
For sake of simplicity, we introduce a notation that corresponds to a discretize semi-norm over the pixels:

$$\left| I_{a,\varepsilon} - Z \circ \Phi_v^1 \right|_{\mathrm{P}}^2 = \sum_{p=1}^{N} \left(I_{a,\varepsilon}(p) - Z \circ \Phi_v^1(p) \right)^2.$$

At last, we define the mean contrast function F given by

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

where $dP(a,\varepsilon)$ is the tensorial product measure on a and ε.

$F(Z)$ must be understood as follows: it measures "on average" how far an image Z is from the image $I_{a,\varepsilon}$ generated from our random warping model using an optimal alignment of Z onto $I_{a,\varepsilon}$. Note that we only observe realizations $I_1, \ldots I_n$ that have been generated with the parameters $a^1, \ldots a^n$ and $\varepsilon^1, \ldots \varepsilon^n$.
However, our goal is to estimate a mean pattern image $Z^\star$ (possibly not unique) which corresponds to the minimum of the contrast function F when $I^\star$ (and of course $dP(a,\varepsilon)$) is unknown. As pointed before, M-estimation will enable is to replace virtually the $\mathbb{E}_{P(a,\varepsilon)}$ by a finite sum $\sum_{i=1}^n$ which depends on the observations and that mimic the former expectation. To estimate $Z^\star$, it is therefore natural to define the following empirical mean contrast:

Definition 3 (Empirical mean contrast)**.** *We define the measure $\mathbb{P}_n$ and the empirical contrast F_n as*

$$\mathbb{P}_n(a,\varepsilon) = \frac{1}{n} \sum_{i=1}^{n} \delta_{a^i,\varepsilon^i} \text{ and } F_n(Z) = \int f(a,\varepsilon,Z) d\mathbb{P}_n(a,\varepsilon).$$

Note that even if we do not observe the deformation parameters a^i and the noise ε^i, it is nevertheless possible to optimize $F_n(Z)$ with respect to Z since it can be written as:

$$F_n(Z) = \frac{1}{n} \sum_{i=1}^{n} \min_{v_i \in \mathcal{V}_A} \left| I_i - Z \circ \Phi_{v_i}^1 \right|_{\mathrm{P}}^2.$$

Moreover, note that in the above equation *it is not required* to specify the law P_A or the law of the additive noise to compute the criterion $F_n(Z)$. We then introduce quite naturally a sequence of sets of estimators

$$\hat{Q}_n = \arg\min_{Z \in \mathcal{Z}} F_n(Z) \tag{6}$$

and we will theoretically compare the asymptotic behavior of these sets with the deterministic one

$$Q_0 = \arg\min_{Z \in \mathcal{Z}} F(Z). \tag{7}$$

In a second time, we will infer in section 3.5 an algorithm to estimate a mean pattern in the set $\hat{Q}_n$. This algorithm consists in a recursive procedure to solve (6). Note that both sets $\hat{Q}_n$ and Q_0 are not necessarily restricted to a singleton.

3.4 Convergence of the estimator

We first state a useful theoretical tool of M-estimation that can be found in Theorem 6.3 of Biscay & Mora (2001). We will use this result to establish the convergence of (6) toward (7). For each integer n, we denote $\hat{Z}_n$ any sequence of images belonging to $\hat{Q}_n$.

Proposition 1. *For any image Z, assume that $F_n(Z) \to F(Z)$ a.s. and that the following two conditions hold*

(C1) the set $\{f(\cdot,\cdot,Z) : Z \in \mathcal{Z}\}$ is an equicontinuous family of functions at each point of $\mathcal{X} = [-A;A]^{2K} \times \mathbb{R}^N$.

(C2) there is a continuous function $\phi : \mathcal{X} \to \mathbb{R}^+$ such that $\int_{\mathcal{X}} \phi(a,\varepsilon)dP(a,\varepsilon) < +\infty$, and for all $(a,\varepsilon) \in \mathcal{X}$ and $Z \in \mathcal{Z}$, $|f(a,\varepsilon,Z)| \le \phi(a,\varepsilon)$.

Then

$$\hat{Q}_\infty \subset Q_0 \ a.s., \tag{8}$$

where $\hat{Q}_\infty$ is defined as the set of accumulation points of the $\hat{Z}_n$, i.e the limits of convergent subsequences $\hat{Z}_{n_k}$ of minimizers $\hat{Z}_n \in \hat{Q}_n$.

The following theorem whose proof can be found in Bigot et al. (2009) gives sufficient conditions to ensure the convergence of the M-estimator in the sense of equation (8).

Theorem 1. *Assume that conditions* **A1** *and* **A2** *hold, then the M-estimator defined through $\hat{Q}_n$ is consistent: any accumulation point of $\hat{Q}_n$ converges to Q_0 almost surely.*

This theorem ensures that the M-estimator, when constrained to live in a fixed compact set of images, converges to a minimizer $Z^\star$ of the limit contrast function $F(Z)$.
Remark that Theorem 1 only proves the consistency of our estimator when the observed images comes from the distribution defined through $I_{a,\varepsilon}$, $(a,\varepsilon) \sim dP(a,\varepsilon)$. This assumption is obviously quite unrealistic, since in practice the observed images generally come from a distribution that is different from the model (4). In Section 4, we therefore address the problem of studying the consistency of our procedure when the observed images $I_i, i = 1,\ldots,n$ are an i.i.d. sample from an unknown distribution on $\mathbb{R}^N$ (see Theorem 2).

3.5 Robustness and penalized estimation

Penalized approach

First, remark that the limite of $\hat{Z}_n$ denoted $Z^\star$ may be equal to the correct pattern $I^\star$ if the observations are generated following the distribution $I_{a,\varepsilon}$, $(a,\varepsilon) \sim dP(a,\varepsilon)$. We address in this paragraph what happens when observations do not follow the law of $I_{a,\varepsilon}$.
In such situation, the minimum Z^* may be very different from the original image I^*, leading to unconsistant estimate as $n \to \infty$. This behaviour is well known in statistics (one may refer to van de Geer (2000) for instance for further details). The loss function has often to be balanced by a penalty which regularizes the matching criterion. In a Bayesian framework, this penalized point of view can be interpreted as a special choice of a prior distribution, e.g Allassonière et al. (2007). In nonparametric statistics, this regularization often takes the form of a penalized criterion which enforces the estimator to belong to a specific space satisfying appropriate regularity conditions.

Image decomposition

It may not be a good thing to just minimize L^2 distance between image and the true warped pattern as the L^2 norm on the euclidean space may not traduce the variability between all the images. Thus, it may be a good point to use another space for images than the set of real vector of size N since this canonical space may not traduce important features of image analysis (edges, textures, etc.) .
In our framework, we point out that choosing to expand the images $Z \in \mathcal{Z}$ into a set of basis functions $(\psi_\lambda)_{\lambda \in \Lambda}$, that are well suited for image processing (e.g. a wavelet basis), is by itself a way to incorporate regularization on $\hat{Z}_n$. Here, the set Λ can be finite or not. More precisely, any image Z can be parametrized by $Z = Z_\theta = \sum_{\lambda \in \Lambda} \theta_\lambda \psi_\lambda$. The estimation of a mean pattern involves the estimation of the coefficients of this image in the basis ψ_l, $\lambda \in \Lambda$. Thus, the penalization term on the image will involve the parameter θ since the basis $(\psi_\lambda)_\lambda$ will remain fixed. This yield the definition of pen_1 as

$$\text{pen}_1(\theta) = \sum_{\lambda \in \Lambda} |\theta_\lambda|.$$

Cost of diffeomorphism

Indeed, the variability may be described in a different way due to the action of the deformations for instance. Thus, in our setting, we may be interested in defining a distance between images that mostly depends on the amount of deformation required to transform the first one to the second with a regular deformation action. A good way to set up such distance is certainly to penalize the energy F with a term $|D^{(m)}\Phi_v^1|$ where m is a derivative of the diffeomorphism. Since each diffeomorphism is parametrized by a finite set of coefficients $(a_k^i)_{i,k}$, the second penalization term of the deformation will concern the unknown a.
Further, we impose regularity on the transformations by adding a penalty term to the matching criterion to exclude unlikely large deformations. The penalty must control both the deformations to avoid too large deformations (see for instance Amit et al. (1991)) and also the images to add a smoothness constraint for the reference image. For this, for Γ a symetric positive definite matrix, define the penalization of the deformation as:

$$\text{pen}_2(v) = \sum_{i=1}^{2} \sum_{k,k'=1}^{K} a_k^i \Gamma_{k,k'} a_{k'}^i.$$

Comments on pen_1 and pen_2

The first penalization term is somewhat classical and corresponds to soft-thresholding estimators, it has been widely used in various context, see for instance the work of Antoniadis & Fan (2001) , and it enables to incorportate some sparsity constraint on the set $\mathcal{Z}$.
For the deformation parameters, the choice $\text{pen}_2(v_a)$ means that we incorporate spatial dependencies using a given matrix Γ. We thus do not assume anymore that all deformations have the same weight, as it was done in the original definition of $F_n(Z)$. Obviously, other choices of penalty can be studied for practical applications and we provide in the sequel consistency results for general penalties. Two parameters λ_1 and λ_2 balance the contribution of the loss function and the penalties. High values of these parameters stress the regularity constraint for the estimator and the deformations.

Finally, we obtain the following estimator $\hat{Z}_n = \sum_{\lambda \in \Lambda} \hat{\theta}_\lambda \psi_\lambda$, with

$$\hat{\theta}_n = \arg\min_{\theta \in \mathbb{R}^\Lambda} \frac{1}{n} \sum_{i=1}^{n} \min_{v_i \in \mathcal{V}_A} \left(\left| I_i - Z_\theta \circ \Phi^1_{v_i} \right|^2_{\mathcal{P}} + \lambda_1 \text{pen}_1(v_i) \right) + \lambda_2 \text{pen}_2(\theta).$$

The effects of adding such extra terms can also be studied from a theoretical point of view. If the smoothing parameters λ_1 and λ_2 are held fixed (they do not depend on n) then it is possible to study the converge of $\hat{\theta}_n$ as n grows to infinity under appropriate conditions on the penalty terms and the set Λ.

More precisely, we address now the problem of studying the consistency of our M-estimator when the observed images (viewed as random vectors in $\mathbb{R}^N$) come from an *unknown* distribution P, that does not necessarily correspond to the model (4). For sake of simplicity we use the notation $\tilde{f}$ introduced in Equation (5), but within a penalized framework with unknown P, the dependency on ε disappears, and $\tilde{f}$ is now defined as

$$\tilde{f}(I, Z_\theta) = \min_{v \in \mathcal{V}_A} \left[\| I - Z_\theta \circ \Phi^1_v \|^2_{\mathcal{P}} + \lambda_1 \text{pen}_1(v) \right] + \lambda_2 \text{pen}_2(\theta), \tag{9}$$

where $\lambda_1, \lambda_2 \in \mathbb{R}^+$, $\text{pen}_1(v) := \text{pen}_1(a) : \mathbb{R}^{2K} \to \mathbb{R}^+$, and $\text{pen}_2(\theta) : \mathbb{R}^\Lambda \to \mathbb{R}^+$. For any θ that "parametrizes" the image Z_θ in the basis $(\psi_\lambda)_{\lambda \in \Lambda}$, let $\tilde{F}$ denote the general contrast function

$$\tilde{F}(Z_\theta) = \int \tilde{f}(I, Z_\theta) dP(I), \tag{10}$$

and $\tilde{F}_n$ the empirical one defined as

$$\tilde{F}_n(Z_\theta) = \frac{1}{n} \sum_{i=1}^{n} \tilde{f}(I_i, Z_\theta).$$

The following theorem, whose proof is deferred to the Appendix, provides sufficient conditions to ensure the consistency of our estimator in the simple case when $\tilde{F}(Z_\theta)$ has a unique minimum at Z_{θ^*} for $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^\Lambda$ is a compact set, and Λ is finite.

Theorem 2. *Assume that Λ is finite, that the set of vector fields $v = v_a \in \mathcal{V}$ is indexed by parameters a which belong to a compact subset of $\mathbb{R}^{2K}$, that $a \mapsto \text{pen}_1(v_a)$ and $\theta \mapsto \text{pen}_2(\theta)$ are continuous. Moreover, assume that $\tilde{F}(Z_\theta)$ has a unique minimum at Z_{θ^*} for $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^\Lambda$ is a compact set. Finally, assume that the basis $(\psi_\lambda)_{\lambda \in \Lambda}$ and the set Θ are such that there exists two positive constants M_1 and M_2 which satisfy for any $\theta \in \Theta$*

$$M_1 \sup_{\lambda \in \Lambda} |\theta_\lambda| \leq \sup_{x \in [0,1]^2} |Z_\theta(x)| \leq M_2 \sup_{\lambda \in \Lambda} |\theta_\lambda|. \tag{11}$$

Then, if P satisfies the following moment condition,

$$\int \|I\|^2_{\infty,N} dP(I) < \infty,$$

where $\|I\|_{\infty,N} = \max_{p=1,\dots,N} |I(p)|$, the M-estimator defined by $\hat{Z}_n = Z_{\hat{\theta}_n}$ where

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \tilde{F}_n(Z_\theta)$$

is consistent for the supremum norm of functions defined on $[0,1]^2$ i.e.

$$\lim_{n \to \infty} \|\hat{Z}_n - Z_{\theta^*}\|_\infty = 0 \quad a.s.$$

4. Experiments and conclusion

4.1 Numerical implementation

Different strategies are discussed in Bigot et al. (2009) to minimize the criterion $\tilde{F}_n(Z_\theta)$. In the numerical experiments proposed in this text, we took the identity matrix for Γ in the formulation of pen_2. We have also chosen to simply expand the images in the pixel basis and have taken $\lambda_1 = 0$ (i.e. no penalization on the space of discrete images viewed as vectors in $\mathbb{R}^N$).

Our estimation procedure obviously depends on the choice of the basis functions $e_k = (e_k^1, e_k^2)$ that generate the vector fields. In the following, we have chosen to use tensor products of one-dimensional B-spline organized in a multiscale fashion. Let $s = 3$ be the order of the B-spline and and $J = 3$. For each scale $j = 0, \ldots, J-1$, we denote by $\phi_{j,\ell}, \ell = 0, \ldots, 2^j - 1$ the 2^j the B-spline functions obtained by taking $2^j + s$ knots points equispaced on $[0,1]$ (see de Boor (1978)). This gives a set of functions organized in a multiscale fashion as shown in Figure 4. Note that as j increases the support of the B-spline decreases which makes them more localized.

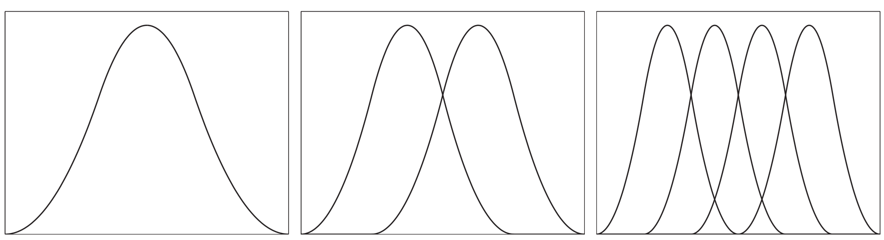


Fig. 4. An example of multiscale B-splines $\phi_{j,\ell}, \ell = 0, \ldots, 2^j - 1$ with $J = 3$ and $s = 3$, ordered left to right, $j = 0, 1, 2$.

For $j = 0, \ldots, J-1$, we then generate a multiscale basis $\phi_{j,\ell_1,\ell_2} : [0,1]^2 \to \mathbb{R},\ \ell_1, \ell_2 = 0, \ldots, J-1$ by taking tensor products the $\phi_{j,\ell}$'s i.e.

$$\phi_{j,\ell_1,\ell_2}(x_1, x_2) = \phi_{j,\ell_1}(x_1)\phi_{j,\ell_2}(x_2).$$

Then, we take $e_k = e_{j,\ell_1,\ell_2} = (\phi_{j,\ell_1,\ell_2}, \phi_{j,\ell_1,\ell_2}) : [0,1]^2 \to \mathbb{R}^2$. This makes a total of $K = \sum_{j=0}^{J-1} 2^{2j} = \frac{2^{2J}-1}{3} = 21$.

Following these choices, one can then use the iterative algorithm based on gradient descent described in Bigot et al. (2009) to find a mean image.

4.2 Mnist database

First we return to the example shown previously in Figure 3 on handwritten digits (Mnist database). As these images are not very noisy, it is reasonable to set $\lambda_1 = 0$ and thus to not use a penalty on the space of images onto which the optimization is done. A value of $\lambda_2 = 10$ to penalize the deformations gave good results.

In Figure 5, we display the naive arithmetic mean Z_{naive} and the mean $\hat{Z}_n$ by minimizing $\tilde{F}_n(Z_\theta)$ obtained from $n = 20$ images of the digits "2". The result obtained with $\hat{Z}_n$ is very satisfactory and is clearly a better representative of the typical shape of the digits "2" in this database then the naive arithmetic mean. Indeed, $\hat{Z}_n$ has sharper edges than the naive mean which is very blurred.

In Figure 6 we finally display the comparison between the naive mean and the mean image $\hat{Z}_n$, for all digits between 0 and 9 with 20 images for each digit. One can see that our approach yields significant improvements. In particular it gives mean digits with sharp edges.

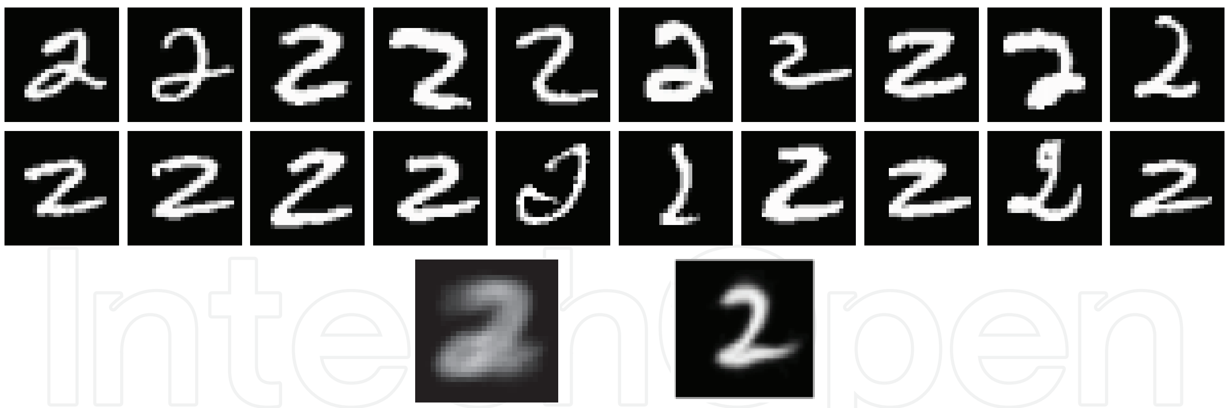


Fig. 5. Naive arithmetic mean (lower left image), mean image $\hat{Z}_n$ (lower right image) based on 20 images of the digit "2" (upper rows).

Fig. 6. Naive arithmetic mean (first row), mean image $\hat{Z}_n$ (second row) based on 20 images from the mnist database.

4.3 Olivetti faces database

Let us now consider a problem of faces averaging. These images are taken taken from the Olivetti face database Samaria & Harter (1994) and their size is $N_1 = 98$ by $N_2 = 112$ pixels. We consider 8 subjects taken from this database. For each subject, $n = 9$ images of the same person have been taken with varying lighting and facial expression. Figure 7 and Figure 8 show the faces used in our simulations.

In Figure 9 and Figure 10 we present the mean images obtained with $\lambda_2 = 1000$, and compare them with the corresponding naive mean. Note that these images are not very noisy, so it is reasonable to set $\lambda_1 = 0$. Obviously our method clearly improves the results given by the naive arithmetic mean. It yields satisfactory average faces especially in the middle of the images.

4.4 Conclusion and perspectives

We have built a very general model of random diffeomorphisms to warp images. This construction relies mainly on the choice of the basis functions e_k for generating the deformations. The choice of the e_k's is relatively large since one is only restricted to take functions with a sufficient number of derivatives that vanish at the boundaries of $[0,1]^2$. Moreover, our estimation procedure does not require the choice of a priori distributions for the random coefficients a_k^i. Other applications of this approach may be developed to obtain some clustering algorithms (K-means adaptation for unsupervised classification) using the energy introduced in this paper. Hence, this model is very flexible as many parameterizations can be chosen. We have only focused on the estimation of the mean pattern of a set of images, but one would like to build other statistics like principal modes of variations of the

Fig. 7. $n = 9$ samples of the Olivetti database for 4 subjects.

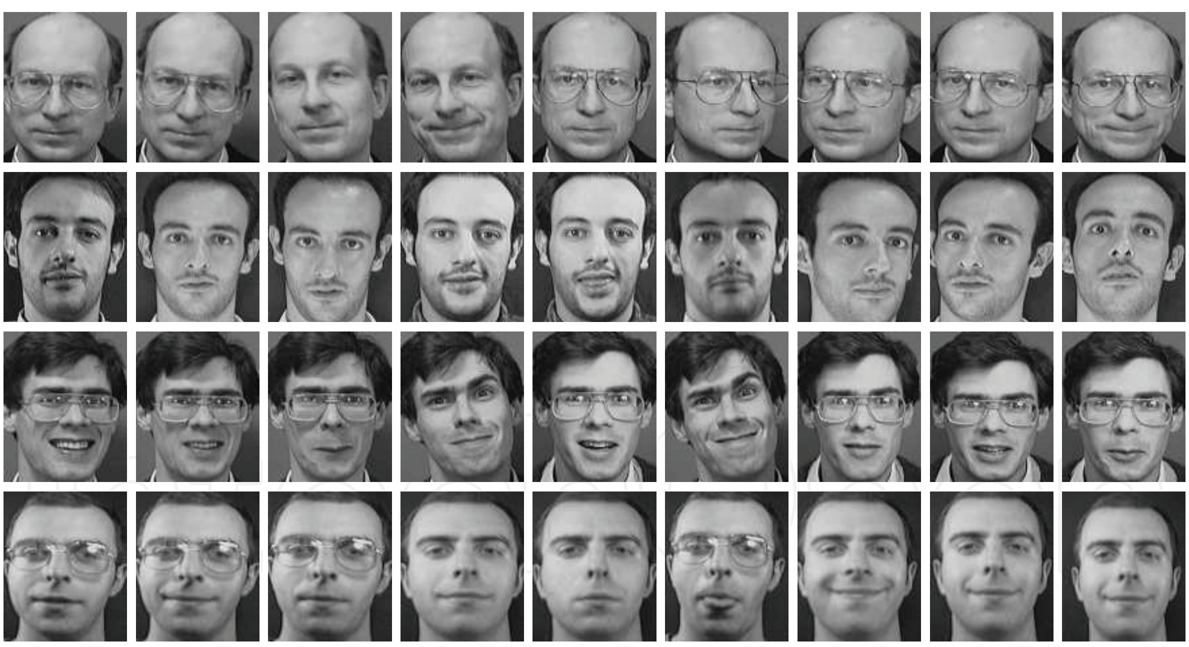


Fig. 8. $n = 9$ samples of the Olivetti database for 4 subjects.

learned distribution of the images or the deformations. Building statistics going beyond the simple mean of set of images within the setting of our model is very challenging for future investigation.

Fig. 9. Example of face averaging for 4 subjects from the Olivetti database. First row: naive arithmetic mean, second row: mean image $\hat{Z}_n$.

Fig. 10. Example of face averaging for 4 subjects from the Olivetti database. First row: naive arithmetic mean, second row: mean image $\hat{Z}_n$.

5. References

Allassonière, S.; Amit, Y. & Trouvé, A. (2007). Toward a coherent statistical framework for dense deformable template estimation, *Journal of the Statistical Royal Society (B)*, Vol. 69, 3-29.

Allassonière, S.; Kuhn, E. & Trouvé, A. (2009). Bayesian Deformable Models building via stochastic approximation algorithm: a convergence study, *Bernoulli*.

Amit, Y.; Grenander, U. & Piccioni, M. (1991). Structural Image Restoration through deformable template, *Journal of the American Statistical Association*, Vol. 86, 376-387.

Antoniadis, A. & Fan, J. (2001). Regularization of Wavelet Approximations, *Journal of the American Statistical Association*, Vol. 96, 939-967.

Beg, M.; Miller, M.; Trouvé, A. & Younes, L. (2005). Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms, *Int. J. Comput. Vision*, Vol. 61, No 2, 139-157.

Bigot, J.; Gadat, S. & Loubes, J.-M. (2009). Statistical M-Estimation and Consistency in large deformable models for Image Warping, *Journal of Mathematical Imaging and Vision*, Vol. 34, No. 3, 270-290.

Biscay, R. & Mora, C. (2001). Metric sample spaces of continuous geometric curves and estimation of their centroids, *Mathematische Nachrichten*, Vol. 229, 15–49.

de Boor, C. (1978). *A practical guide to splines*, Applied Mathematical Sciences, Springer-Verlag, New York, pp. 368.

Candes, E. & Donoho, D. (2000). Recovering Edges in Ill-Posed Inverse Problems: Optimality of Curvelet Frames, *Annals of Statistics*, Vol. 30, 784-842.

Faugeras, O. & Hermosillo, G. (2002). Well-posedness of eight problems of multi-modal statistical image-matching, *Biomedical Imaging*, Vol. 15, No. 23, 1-64.

Glasbey, C. & Mardia, K. (2001). A penalized likelihood approach to image warping, *Journal of the Statistical Royal Society (B)*, Vol. 63, No. 3, 465-514.

Grenander, U. (1993). *General pattern theory - A mathematical study of regular structures*, Clarendon Press.

Grenander, U. & Miller, M. (2007). *Pattern Theory: From Representation to Inference*, Oxford Univ. Press.

Huilling, L. (1998). On the consitency of Procrustean mean Shapes, *Advances in Applied Probability*, Vol. 30, 53-63.

Joshi S.; Davis B.; Jomier M. & Gerig G. (2004). Unbiased diffeomorphic atlas construction for computational anatomy, *Neuroimage*, vol. 23, 151-160.

Le Cun, Y.; Bottou, L.; Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, Vol. 86, No. 11, 2278-2324.

Ma, J.; Miller, M.; Trouvé, A. & Younes, L. (2008).Bayesian Template Estimation in Computational Anatomy, *NeuroImage*, Vol. 42, No. 1, 252-26.

Miller M. & Younes L. (2001). Group Actions, Homeomorphisms, And Matching: A General Framework,*International Journal of Computer Vision*, vol. 41, 61-84.

Samaria, F.S. & Harter, A. (1994). Parameterisation of a Stochastic Model for Human Face Identification, *WACV94*, 138-142.

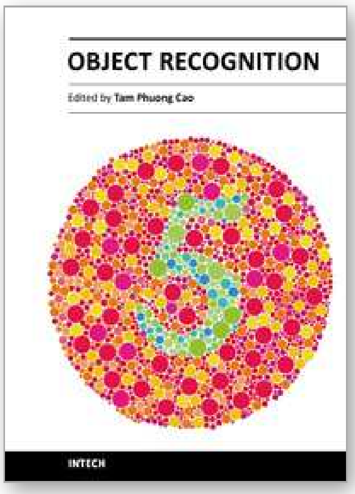
Trouvé, A. & Younes, L. (2005a). Local geometry of deformable templates, *SIAM Journal on Mathematical Analysis*, Vol. 37, No. 1, 17-59.

Trouvé, A. & Younes, L. (2005b). Metamorphoses Through Lie Group Action, *Foundations of Computational Mathematics*, Vol. 5, No 2, 173-198.

van de Geer, S. (2000). *Applications of empirical process theory*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, xii+286.

van der Waart, A. (1998). *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics 03, Cambridge Univ. Press , New-York, pp. 460.

Younes, L. (2004). *Invariance, déformations et reconnaissance de formes*, Mathematics & Applications, Springer-Verlag, Berlin, pp. xviii+248.



Object Recognition
Edited by Dr. Tam Phuong Cao

ISBN 978-953-307-222-7
Hard cover, 350 pages
Publisher InTech
Published online 01, April, 2011
Published in print edition April, 2011

Vision-based object recognition tasks are very familiar in our everyday activities, such as driving our car in the correct lane. We do these tasks effortlessly in real-time. In the last decades, with the advancement of computer technology, researchers and application developers are trying to mimic the humanâ€™s capability of visually recognising. Such capability will allow machine to free human from boring or dangerous jobs.