Metropolis-Hasting techniques for finite element-based registration

Frederic J.P. Richard and Adeline Samson
MAP5 CNRS UMR 8145, University Paris Descartes.
[samson,richard]@math-info.univ-paris5.fr

Abstract

In this paper, we focus on the design of Markov Chain Monte Carlo techniques in a statistical registration framework based on finite element basis (FE). Due to the use of FE basis, this framework has specific features. The main feature is that displacement random fields are markovian. We construct two hybrid Gibbs/Metropolis-Hasting algorithms which take fully advantage of this markovian property. The second technique is defined in a coarse-to-fine way by introducing a penalization on the sampled posterior distribution. We present some promising results suggesting that both techniques can accurately register images. Experiments also show that the penalized technique is more robust to local maxima of the posterior distribution than the first technique. This study is a preliminary step towards the estimation of model parameters in complex image registration problems.

Note: This preprint is published in the CVPR Workshop on Image Registration and Fusion, June 23, 2007; Minneapolis, Minnesota, USA

1. Introduction

Researchs on Image Registration are among the most active of Image Processing (see [1] for a review). Since early works at the end of 80's, numerous models have been proposed to deal with a wide variety of registration problems encountered in practice (large deformations, multimodality registration, class-dependant registration, template registration, ...). There is however very few literature on how to select and adapt registration models to data of a specific application. Model adaptation and selection is a crucial problem which is still open.

In [2], Glasbey and Mardia set up a stochastic framework based on a penalized likelihood approach in which some model parameters (penalization weight, template) are adapted to data by a statistical estimation. More recently, Allassonniere and colleagues propose a Bayesian statistical framework in which they define a consistent estimator of model parameters (stiffness matrix, template, image noise variance) [3, 4]. In [3], the estimation technique is based on a stochastic version of the EM algorithm [5] known as SAEM (Stochastic Approximation Expectation Maximization) in the Statistics community. SAEM algorithm in [3] contains a simulation step in which displacements are sampled from a probability distribution using a Markov Chain Monte Carlo technique (MCMC). MCMC techniques are used because the probability distribution does not have an analytic form and cannot be sampled directly. The design of MCMC in SAEM is a critical issue, especially when considering highly complex image registration problems.

In this paper, we focus on this issue. We present first investigations into a particular statistical framework based on finite element basis (FE). Due to the use of FE basis, this particular framework has several specific features which we enlighten and take advantage of. The main feature of the FE-based framework is that displacement random fields are gibbsian and markovian. This is also the case in the framework based on finite differences in [6], but not for common frameworks based on spectral decompositions [7].

The markovian property enables to define MCMC techniques by combining Metropolis-Hastings (MH) and Gibbs procedures. We construct two hybrid MH/Gibbs techniques and apply them to the registration problem defined as a Maximum A Posteriori estimation. We test and compare the two techniques on some synthetic data.

2. Variational framework

2.1. Problem Formulation

In practice, bidimensional images are observed on regular grids $\mathcal{G} = \{1, \cdots, N_1\} \times \{1, \cdots, N_2\}$ of size $n_g = N_1 N_2$. Any observed image y is thus described by a set of gray-level values y_n at positions $n = (n_1, n_2)$ of $\mathcal{G}$. For deforming an image y, it is however necessary to represent it as a function defined on a connected domain. Hence we extend the definition of any discrete image y on the domain $\mathcal{D} = [0, N_1] \times [0, N_2]$, by setting $f(1/2 + n_1, 1/2 + n_2) =$

y_n for all $n = (n_1, n_2)$ of $\mathcal{G}$ and interpolating f on the whole domain $\mathcal{D}$. Here, we use a bilinear interpolation technique and have continuous images f.

An image f defined on $\mathcal{D}$ is registered to another image g using coordinate changes. A coordinate change ϕ is a continuous and invertible function (an homeomorphism) mapping the image domain $\mathcal{D}$ onto itself. A coordinate change ϕ induces a geometrical deformation f_ϕ of an image f which is defined as $f_\phi = f \circ \phi$. Coordinate changes ϕ are also called deformations and can be decomposed as $\phi = \mathrm{id} + u$, where id is the identity function and u are displacements. The registration problem consists of finding a map ϕ which makes the deformed image f_ϕ and the target image g as similar as possible. This is usually formulated in terms of an inverse problem: find a displacement u in a functional space $\mathcal{V}$ which minimizes an energy of the form

$$\mathcal{E}(u) = \mathcal{S}(f_{\mathrm{id}+u}, g) + \mathcal{R}(u). \tag{1}$$

The energy is composed of two terms. The first term $\mathcal{S}$ defines the image similarity criterion and is as low as the deformed image $f_{\mathrm{id}+u}$ and the target image g are similar. Here, $\mathcal{S}$ is the usual distance on $L^2(\mathcal{D})$

$$\mathcal{S}(f, g) = \frac{1}{2}\int_{\mathcal{D}} \left(f(x) - g(x)\right)^2 dx = \frac{1}{2}|f - g|^2_{\mathcal{D}}. \tag{2}$$

The second term is a regularization term which ensures that the problem is well-posed in the space $\mathcal{V}$. It is usually of the form $\mathcal{R}(u) = \frac{1}{2}a(u, u)$, where $a(\cdot, \cdot)$ is a bilinear form on $\mathcal{V}$. In this paper, $\mathcal{R}$ is the strain energy of linearized Elasticity, defined as

$$\mathcal{R}(u) = \frac{1}{2}\int_{\mathcal{D}} \lambda \mathrm{tr}^2(e(u)) + 2\mu\, \mathrm{tr}(e(u)^T\, e(u)) dx, \tag{3}$$

where λ and μ are Lame constants and $e(u)$ is the linearized strain tensor $e(u) = \frac{1}{2}(\nabla u^T + \nabla u)$. We set $\mathcal{V} = H^1(\mathcal{D}) \times H^1(\mathcal{D})$, where $H^1(\mathcal{D})$ is an usual Sobolev space on $\mathcal{D}$ and assume homogeneous Dirichlet boundary conditions ($u(x) = 0$ on the boundary of $\mathcal{D}$).

The minimization problem described by Equation (1) is commonly solved using a Riesz technique. Such a technique consists of defining approximate problems in some finite-dimensional subspaces $\mathcal{V}^h$ of $\mathcal{V}$ spanned by a family of elements of $\mathcal{V}$. Spectral techniques [7] and finite element methods are two different kinds of the Riesz method.

2.2. Finite Elements

The finite element method enables to construct approximation subspaces $\mathcal{V}^h$ [8]. The construction of $\mathcal{V}^h$ is fundamentally based on a triangulation of the domain $\mathcal{D}$. A triangulation of the domain $\mathcal{D}$ is a set $\mathcal{T}^h$ composed of non-overlapping triangular or rectangular subregions of $\mathcal{D}$ (called finite elements) which cover the whole domain $\mathcal{D}$. Deformation spaces $\mathcal{V}^h$ built with the finite element method have several specific properties: (i) the restriction of any element of $\mathcal{V}^h$ to a finite element is polynomial, (ii) elements of $\mathcal{V}^h$ are continuous (or even continuously differentiable), (iii) there exists a finite-dimensional basis of $\mathcal{V}^h$ composed of functions which have small supports limited to a few finite elements.

We now give a construction of a space $\mathcal{V}^h$ using P_1-Lagrange rectangular finite elements. Let us fix a finite element size h dividing N_1 and N_2. The domain $\mathcal{D}$ is triangulated by rectangular finite elements $t_e = [e_1, he_1] \times [e_2, he_2]$ defined for indices $e = (e_1, e_2)$ of the set $\mathcal{K}^h = \{0, \cdots, N_1/h - 1\} \times \{0, \cdots, N_2/h - 1\}$. We order vertices of all finite elements in an index set $\mathcal{N}^h$ of size n_v. Given a vertex s_k for k in $\mathcal{N}^h$, we denote by $\mathcal{T}_k = \{e \in \mathcal{K}^h, s_k \in t_e\}$ the index set of all finite elements it belongs to. Given a finite element t_e, we denote by $\mathcal{O}_e = \{k \in \mathcal{N}^h, s_k \in t_e\}$, the index set of all vertices which belongs to t_e. We denote by $v_k = (v_{k1}, v_{k2})^T$ the displacement vector at a vertex position s_k.

We state that the components of restrictions $u_{|t_e}$ of any function u of $\mathcal{V}^h$ on any element t_e are both bilinear, i.e. of the form $a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2$. It is well-known that restrictions $u_{|t_e}$ of u on t_e are characterized by values of u at the vertices of t_e. Furthermore, the restriction $u_{|t_e}$ can be expressed as $u_{|t_e}(x) = \sum_{k \in \mathcal{O}_e} \sum_{i=1}^{2} v_{ki}\; u^e_{ki}(x)$, where the u^e_{ki} are eight local basis functions (called shape functions) having support on t_e. On the whole domain $\mathcal{D}$, displacements u of $\mathcal{V}^h$ can be expressed in a basis of $n_h = 2n_v$ functions. Each of these functions are associated to a vertex s_k and are defined for all $k \in \mathcal{N}^h$ and $i \in \{1, 2\}$ from the local basis functions as $u^h_{ki} = \sum_{e \in \mathcal{T}_k} u^e_{ki}$. We have

$$\mathcal{V}^h = \{u(x) = \sum_{k \in \mathcal{N}^h} \sum_{i=1}^{2} v_{ki}\; u^h_{ki}(x), v \in \mathbb{R}^{2n_v}\}. \tag{4}$$

The elements of $\mathcal{V}^h$ are continuous on $\mathcal{D}$. When expressed in the subspace $\mathcal{V}^h$, the minimization problem becomes: find a coefficient vector $v = (v_k)_{k \in \mathcal{N}^h}$ in $\mathbb{R}^{n_h}$ which minimizes a discrete energy of the form

$$\mathcal{E}^h(v) = \mathcal{R}^h(v) + \mathcal{S}^h(v), \tag{5}$$

where $\mathcal{R}^h$ and $\mathcal{S}^h$ are discrete versions of $\mathcal{R}$ and $\mathcal{S}$

$$\mathcal{R}^h(v) = \frac{1}{2}v^T\Gamma^h v \quad \text{and} \quad \mathcal{S}^h(v) = \frac{1}{2}|f_v - g|^2_{\mathcal{D}}. \tag{6}$$

In this expression, f_v denotes the deformed image $f(\mathrm{id} + \sum_{k \in \mathcal{N}^h} \sum_{i=1}^{2} v_{ki}\; u^h_{ki})$ and Γ^h is a $n_h \times n_h$-matrix (called the stiffness matrix) whose elements $\Gamma^h_{(ki;lj)} = a(u^h_{ki}, u^h_{lj})$ for all $k, l \in \mathcal{N}^h$ and $i, j \in \{1, 2\}$.

Besides, the support of the basis function u^h_{ki} is restricted to the finite elements which the vertex s_k belongs to (i.e.

$\cup_{e\in\mathcal{T}_k} t_e$). Hence, the inner product $a(u^h_{ki}, u^h_{lj}) = 0$ whenever the vertices s_k and s_l are not on a common finite element. As a consequence, the stiffness matrix Γ^h is sparse and the stiffness term $\mathcal{R}^h(v)$ in the energy $\mathcal{E}^h(v)$ can be simplified as follows

$$\frac{1}{2}\sum_{k\in\mathcal{N}^h}\sum_{i,j=1}^{2} v_{ki}\Gamma^h_{(ki;kj)}v_{kj} + \sum_{k\in\mathcal{N}^h}\sum_{l\in\mathcal{N}_k}\sum_{i,j=1}^{2} v_{ki}\Gamma^h_{(ki;lj)}v_{lj},$$

where $\mathcal{N}_k$ denotes the set of vertices which are different from s_k and on a same finite element as s_k. Furthermore, the similarity term $\mathcal{S}^h(v)$ in Equation (5) can be decomposed as a sum of local integrals on finite elements

$$\frac{1}{2}\sum_{e\in\mathcal{K}^h}\int_{t_e}(f(x+\sum_{i=1}^{2}\sum_{k\in\mathcal{O}_e} v_{ki}\, u^e_{ki}(x)) - g(x))^2 dx.$$

The similarity energy can also be written in a discrete form $\mathcal{S}^h_d(v)$ as

$$\frac{1}{2}\sum_{e\in\mathcal{K}^h}\sum_{n\in\mathcal{G},x_n\in t_e}(f(x_n+\sum_{i=1}^{2}\sum_{k\in\mathcal{O}_e} v_{ki}\, u^e_{ki}(x_n)) - y_n)^2,$$

where $x_n = (n_1 + 1/2, n_2 + 1/2)$ and $y_n = g(x_n)$.

3. Stochastic Framework

3.1. Problem Formulation

In the remaining of the text, we will denote random variables or fields with upper cases and deterministic components or random realizations with lower cases. An usual stochastic representation of deformations can be obtained using displacement decompositions in subspaces $\mathcal{V}^h$. Let $\mathcal{V}^h$ be defined as in Section 2.2 and $V = (V_k)_{k\in\mathcal{N}^h}$ be a centered Gaussian vector with covariance matrix Σ. For any fixed position x of $\mathcal{D}$, we define a random vector $U(x) = \sum_{k\in\mathcal{N}^h}\sum_{i=1}^{2} V_{ki}u^h_{ki}(x)$. The collection $U = (U(x), x\in\mathcal{D})$ is a Gaussian vector-valued random field on $\mathcal{D}$ which is characterized by a covariance structure depending only of the matrix Σ and elements u^h_{ki}. The distribution law of the Gaussian vector V will be called the prior deformation distribution and denoted by $\pi(v;\Sigma)$. An observed image can be represented as a collection $Y = \{Y_n, n\in\mathcal{G}\}$ of random variables defined on all sites of the grid. Such a collection is as a random field on $\mathcal{G}$. We assume that the target image Y is a random field on $\mathcal{G}$ and that it depends statistically on deformations and a fixed image f defined on $\mathcal{D}$. This dependency is defined through the following statistical model

$$\forall\, n\in\mathcal{G},\ Y_n = f(x_n + \sum_{k\in\mathcal{N}^h}\sum_{i=1}^{2} V_{ki}u^h_{ki}(x_n)) + B_n, \quad (7)$$

where the B_n are n_g independant centered Gaussian variables with variance σ^2. The density function representing the distribution of image observations given deformations and a fixed image f is called the conditional density and denoted by $p(y|V; f, \sigma)$. The density function of displacements given images is called the posterior distribution and denoted by $p(v|y; f, \sigma, \Sigma)$. This distribution can be derived from the prior deformation distribution and the conditional distribution using a Bayes formulae

$$p(v|y; f, \sigma, \Sigma) = \frac{p(y|v; f, \sigma)\pi(v; \Sigma)}{\int p(y|V = w; f, \sigma)\pi(w; \Sigma)dw}. \quad (8)$$

In this framework, the registration problem can be formulated in terms of a Maximum A Posteriori (MAP or posterior mode) problem [7, 2]: given an observed image y and a reference image f, find a deformation $\hat{v}_{\mathrm{MAP}}$ which maximizes the posterior density distribution $p(v|y; f, \sigma, \Sigma)$.

Now, consider a discretization $\mathcal{E}^h_d$ of the energy $\mathcal{E}^h$ in Equation (5)

$$\frac{1}{2}v^T\Gamma^h v + \frac{1}{2}\sum_{n\in\mathcal{G}}(f(x_n + \sum_{k\in\mathcal{N}^h}\sum_{i=1}^{2} V_{ki}u^h_{ki}(x_n)) - y_n)^2.$$

If we set $\Gamma^h = \Sigma^{-1}$ and $\sigma = 1$, we can notice that $-\log(p(v|y; f, \sigma, \Sigma)) = \mathcal{E}^h_d(v) + c(\Sigma)$, where $c(\Sigma)$ is a constant depending only on Σ. Hence the minimization problem defined by Equation (5) and the MAP problem can be considered as two similar problems.

In the remaining of the paper, we take the posterior density function defined from discrete energies $\mathcal{R}^h$ and $\mathcal{S}^h_d$ in Equation (6) as

$$p(v|y; f, \sigma, \Sigma) = \frac{1}{Z}\exp(-\mathcal{R}^h(v) - \frac{1}{\sigma^2}\mathcal{S}^h_d(v)), \quad (9)$$

where $\Sigma^{-1} = \Gamma^h$ and Z is a normalization constant depending on σ and Σ. In the next section, we focus on the properties of local displacement interactions on vertices.

3.2. Markov Random Field definition

We define a neighbor system on the vertex set $\mathcal{N}^h$: we state that two vertices s_k and s_l are neighbors if and only if $k \neq l$ and s_k and s_l are on a common finite element t_e. The index set of neighbors of s_k corresponds to the set denoted $\mathcal{N}_k$ in Section 2.2. The set $\mathcal{N}^h$ equipped with the neighbor system is a graph. Cliques of the graph are subsets of $\mathcal{N}^h$ which are either singletons (single elements of $\mathcal{N}^h$) or composed of pairwise-neighbor elements. In the graph, clique orders are below four, since there are only four vertices defining a finite element. Cliques of order four are the sets formed by vertices of a finite element t_e. These sets are denoted $\mathcal{O}_e$ in Section 2.2.

We notice that the posterior distribution defined by Equation (9) is of the form

$$\frac{1}{C}\exp\left(\sum_{k\in\mathcal{N}^h}\mathcal{P}_k(v)+\sum_{k\in\mathcal{N}^h}\sum_{l\in\mathcal{N}_k}\mathcal{P}_{kl}(v)+\sum_{e\in\mathcal{K}^h}\mathcal{P}_e(v)\right),$$

where potentials $\mathcal{P}_k(v) = -\frac{1}{2}\sum_{i,j=1}^{2} v_{ki}\Gamma^h_{(ki;lj)}v_{kj}$ depend only on values of v on a singleton k, potentials $\mathcal{P}_{kl}(v) = -\sum_{i,j=1}^{2} v_{ki}\Gamma^h_{(ki;lj)}v_{lj}$ on values on a clique of order 2 and potentials $\mathcal{P}_e(v) = -\frac{1}{2\sigma^2}\sum_{n\in\mathcal{G},x_n\in t_e}(f(x_n + \sum_i\sum_{k\in\mathcal{O}_e} v_{ki}\,u^e_{ki}(x_n)) - y_n)^2$ on values on a clique of order 4. In other words, the posterior density has a Gibbs form (with respect to the neighbor system). Hence, according to the Hammersley-Clifford theorem [9], the field V given Y is a Markov field with respect to the neighbor system. This means that displacements on a vertex depend only on those on neighbor vertices and that laws governing displacement interactions on vertices are local. The Markovian property of the field directly results from the fact that shape functions u^h_{ki} have small supports and that the stiffness matrix Γ is sparse. In [7], shape functions derived from spectral decomposition have not a local support and consequently, the similarity term cannot be defined as a sum of local potentials.

4. MAP estimation

4.1. MCMC techniques

Due to the Bayes formula in Equation (8), solving the MAP problem is equivalent to finding a deformation maximizing $p(y|V = v; f, \sigma)\pi(v; \Sigma)$ with respect to v. When the posterior distribution does not have a simple analytic form, finding its mode may be difficult or even impossible. In such cases, one can alternatively simulate a sample $(v^{(t)})_{t=1,\dots,T}$ from the posterior distribution using Markov Chain Monte Carlo (MCMC) techniques and then estimate the MAP as the sample mode. The MCMC technique principle consists of simulating an ergodic Markov chain with stationary distribution $p(v|y; f, \sigma, \Sigma)$ [10, 11]. The most popular MCMC techniques are the iterative Gibbs sampler and the Metropolis-Hastings (MH) algorithm, which ensure the existence of a transition kernel such that the stationary distribution of the generated Markov chain is the target distribution $p(v|y; f, \sigma, \Sigma)$.

The Metropolis-Hastings algorithm is commonly used when it is not possible to sample directly from the conditional probability. The MH algorithm is fundamentally based on a so-called instrumental distribution q_t from which it is possible to sample. In our MAP context, a MH algorithm will typically proceed as follows.

1. Initialization of $v^{(0)}$.
2. Iteration $t, t \geq 1$.
   - Sample $v' \sim q_t(v^{(t-1)}, v')$.
   - Evaluate the acceptance probability

$$\rho_t(v^{(t-1)}, v') = \min\left(\frac{p(v'|y)}{p(v^{(t-1)}|y)}\frac{q_t(v', v^{(t-1)})}{q_t(v^{(t-1)}, v')}, 1\right).$$

   - Sample a realization u of an uniform law on $[0, 1]$. If $u \leq \rho_t(v^{(t-1)}, v')$ then set $v^{(t)} = v'$, otherwise set $v^{(t)} = v^{(t-1)}$.
3. Set $t = t + 1$ and go to step 2).

The choice of the instrumental distribution q in MH algorithm is a critical issue. Although the convergence of the MH algorithm is ensured under some generic assumptions on q [10], the rate of convergence depends strongly on q and can be very slow. Intuitively, an instrumental distribution q will be efficient whenever it samples candidates in regions where the posterior density is the highest.

One could simply choose an instrumental distribution q which changes v' globally on all vertex points. The two classical instrumental distributions are (i) $q(v'; v) = \pi(v'; \Sigma)$, the prior deformation distribution, leading to an independent MH algorithm, and (ii) $q(v'; v)$ such that $v' = v + \eta$ where η is sampled from centered multivariate Gaussian distribution (random walk). The random walk scheme which samples new candidate v' depending on current displacements v, is known to be more efficient than independent MH algorithms. However both approaches are not relevant in our context. Indeed, due to the large number of vertex points, the chance to accept a new candidate is very low. Consequently, the Markov chain converges very slowly towards its stationary distribution. Moreover, whenever the posterior distribution is higher for a new displacement candidate v' than for current displacements v, displacements are completely updated without taking into account possible values v_k of v which are more likely than v'_k from a local viewpoint. For these reasons, we believe it is preferable to sample new candidates v' which differ from v only on a few vertices. Hence, in the next section we propose a local version of the MH algorithm which is based on a Gibbs sampler. Here, the use of a Gibbs sampler is feasible and efficient due to the markovian properties of random fields.

Besides, the convergence of MCMC techniques strongly depends on the initialization if the posterior distribution $p(v|y; f, \sigma, \Sigma)$ has several local maxima. In such cases, similarly to a simulated annealing scheme [9], a penalized MH algorithm (PMH) can improve the convergence towards the global maximum. Such an algorithm is based on a *penalization* $\beta_t > 0$ depending on time t. A PMH is a simulation of a non-homogeneous Markov chain with a target penalization-dependent distribution $p_{\beta_t}(v|y; f, \sigma, \Sigma)$. In

PMH, the choice of the initial penalization β_0 and the sequence of penalization is a critical issue. The definition of the penalized distribution is also important. All these points will be discussed in the next section.

4.2. Hybrid Gibbs-MH algorithms

In this paper, we use a Gibbs sampler-random walk scheme for sampling new candidates. Let s_k be a fixed vertex ($k \in \mathcal{N}^h$), a new candidate is sampled by setting $v'_l = v_l$ for $l \neq k$ and $v'_k = v_k + \eta$, where $\eta = (\eta_1, \eta_2)^T$ are two independent realizations of a centered Gaussian random variable with standard deviation $\delta_t > 0$. We denote by $q_k^t(v, v')$ the instrumental law corresponding to this scheme at iteration t. This law is symmetrical ($q_k^t(v, v') = q_k^t(v', v)$). We define a Gibbs sampler based on this random-walk scheme.

1. Initialization. Set $v^{(0)} = (0)$.
2. Iteration $t, t \geq 1$. For all vertex indices $k \in \mathcal{N}^h$,
   - Define the current displacement w as $w_l = v_l^{(t)}, l < k$ and $w_l = v_l^{(t-1)}, l \geq k$,
   - Sample $v' \sim q_k^t(w, v')$.
   - Evaluate the acceptance probability
   $$\rho_t(w, v') = \min\left(\frac{p_{\beta_t}(v'|y)}{p_{\beta_t}(w|y)}, 1\right).$$
   - Sample a realization u of an uniform law on $[0, 1]$. If $u \leq \rho_t(w, v')$ then set $v_k^{(t)} = v'_k$, otherwise set $v_k^{(t)} = v_k^{(t-1)}$.
3. Set $t = t + 1$ and go to step 2).

Due to markovian properties, the computation of the acceptance probability $\rho_t(w, v')$ depends only on the displacement values of the vertex k and its neighbors.

We implemented two versions of this algorithm which differ in the choice of sequences of the scaling parameter δ_t and the penalization β_t. In the first version (refer to as MH, for Metropolis-Hasting algorithm), scaling parameter and penalization are constant for all iterations $t > 0$. In the second version (refer to as PMH, for penalized Metropolis-Hastings algorithm), the scaling parameter and the penalization vary with iterations.

In PMH, the scaling parameter and penalization sequences are defined in such a way that (i) during first iterations the algorithm attempts coarse image difference corrections with large displacements and (ii) as iterations increase, it tries more and more accurate corrections. Such an approach is analogous to the coarse-to-fine strategies which are commonly implemented in gradient descent algorithms. According to this strategy, the scaling parameter sequence is defined so that large displacements η are proposed during first iterations and as iterations increase, displacements become smaller and smaller: $\delta_t = \delta_{max}\tau_t + \delta_{min}(1 - \tau_t)$, where δ_{min} and δ_{max} are lower and upper bounds for δ_t variations and $\tau_t = (0.985)^t$. The sequence of τ_t is chosen so that the algorithm spends more time in correcting details than in making coarse corrections. We tried a linear sequence $\tau_t = (T_{\max} - t)/T_{\max}$ and obtained worse results. The values of δ_{min} and δ_{max} can be interpreted as the lowest expected precision and the largest expected displacements, respectively (in experiments, $\delta_{min} = 1$ and $\delta_{max} = 30$). For the penalization, we choose a decreasing sequence $(\beta_t)_{t\geq 0}$ of the form $\beta_t = 1/(1 - 0.999\tau_t)$. At iteration t, the penalized posterior distribution is defined as

$$p_{\beta_t}(v|y; f, \sigma, \Sigma) \propto p(y|V = v; f, \sigma)(\pi(v; \Sigma))^{\beta_t}.$$

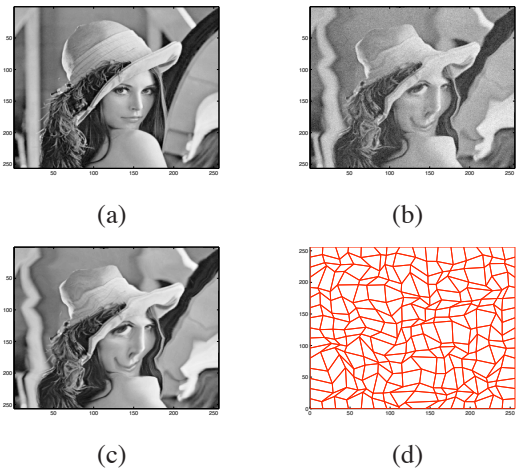
Notice that the penalization is expressed on the regularization term $\pi(v; \Sigma)$ and enforces regularization constraints. Consequently, during first iterations, candidate displacements which are simulated through large displacements, are accepted only if they generate significant image difference corrections (coarse corrections). As iterations increase, the regularization constraint is relaxed and candidate displacements producing small corrections become more easily accepted.

5. Results

For an evaluation, we simulated hundred deformations of the image Lena. For that, we first sampled deformations from the prior distribution using a Gibbs sampler (values of Lame constants were $\lambda = 10^{-6}$ and $\mu = 0.005$ which correspond to a Young modulus $E = 10^{-2}$ and a Poisson ratio $\nu = 10^{-5}$, finite elements were squares of size 16×16 pixels). We then simulated deformed images by applying deformations to Lena and adding a centered Gaussian noise with standard deviation $\sigma = 10$. A deformation example is shown on Figure 1 (b).

We applied simple and penalized Metropolis-Hasting algorithms (respectively MH and PMH) described in Section 4.2 to the deformation of the original image Lena onto each of the simulated images (both algorithms were applied with $T = 400$ iterations). This application is illustrated on Figure 1. After each registration, we computed a mean displacement error (MDE) by averaging the euclidean distances between the estimated displacements and the true displacements at node points. We also computed the standard deviation of the image differences after registration (SDD).

In terms of MDE, MH and PMH were not significantly different. Averaged over all the simulated cases, MDE of MH was 1.16 ± 2.67 pixels whereas MDE of PMH was 1.06 ± 2.35 pixels. MH produced worse MDE than PMH in



(a) (b) (c) (d)

Figure 1. (a) original source image Lena; (b) simulated target image; (c) deformation of the source image onto the target image estimated by Penalized Metropolis-Hastings; (d) deformed mesh.

60 percent of the cases. From the SDD viewpoint, MH and PMH performances were more different. On average, SDD of MH and PMH were 12.2 and 11, respectively. Moreover, MH had higher SDD than PMH in 82 percent of the cases. On Figure 2, we plotted SDD values of PMH vs SDD values of MH for each of the simulated cases. SDD values of PMH are mostly concentrated just above the value of the noise standard deviation ($\sigma = 10$) whereas values of MH are quite uniformly spread on the interval $[10, 14]$. There are also many cases where the performance of MH is low (SDD$>$ 14) whereas performances of PMH is reasonable (SDD$<$ 12). This corresponds to cases where PMH avoids some local maxima in which MH is stopped.

6. Conclusion

We defined a statistical image registration framework in which deformation random fields are markovian. In this framework, we constructed two hybrid Gibbs/Metropolis-Hasting techniques which take fully advantage of these markovian properties. The second technique was defined in a coarse-to-fine way by introducing a penalization on the sampled posterior distribution. We presented some promising results suggesting that both techniques can accurately register images. Experiments also showed that the penalized technique is more robust to local maxima of the posterior distribution than the first technique. This study is the preliminary step towards the development of estimation techniques which could be applied to complex image registration problems. In the future, we will integrate these techniques in a SAEM parameter estimation method.

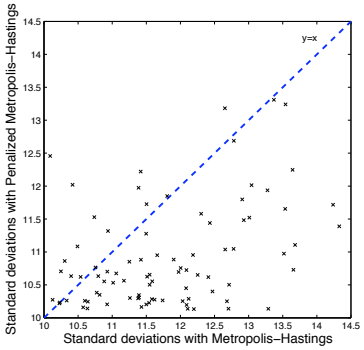


Figure 2. Standard deviation values of the image differences obtained with PMH (ordinate axis) versus values obtained with MH (abscissa axis) (100 experiments).

References

[1] J. Maintz and M. Viergever. A survey of medical image registration. *Med. Im. Anal.*, 2(1):1–36, 1998. 1

[2] C.A. Glasbey and K.V. Mardia. A penalized likelihood approach to image warping. *J. R. Stat. Soc. (Series C)*, 63(3):465–514, 2001. 1, 3

[3] S. Allassonniere, E. Kuhn, A. Trouve, and Y. Amit. Generative model and consistent estimation algorithms for non-rigid deformable models. In *Proc. of IEEE ICASSP*, Toulouse, France, 2006. 1

[4] S. Allassonniere, Y. Amit, and A. Trouve. Towards a coherent statistical framework for dense deformable template estimation. *J. R. Stat. Soc.: Series B (Statistical Methodology)*, 69(1):3–29, Feb 2007. 1

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B*, 39:1–38, 1977. 1

[6] F. Richard. A comparative study of markovian and variational image-matching techniques in application to mammograms. *Patt. Recog. Let.*, 26(12):1819–1829, 2005. 1

[7] Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable templates. *J. Am. Stat. Assoc.*, 86(414):376–387, 1991. 1, 2, 3, 4

[8] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland, Amsterdam, 1978. 2

[9] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE PAMI*, 6:721–741, 1984. 4

[10] L. Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1762, 1994. 4

[11] C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer-Verlag, 2002. 4